



# 視覺化的語言心法

## 資料、模型、與溝通

# 關於我

- › 呂奕 LEO LU
- › 台大工管
- › 目前於金融業服務
- › BUILD DATA PRODUCTS
  - › ETL
  - › MODELS
  - › TEXT MINING
  - › VIZ
- › ...



# 大部分人對視覺化的常見問題

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難
- › 覺得視覺化需要天生的設計美感？

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難
- › 覺得視覺化需要天生的設計美感？
  - › 好的視覺化是不是主觀的？

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難
- › 覺得視覺化需要天生的設計美感？
  - › 好的視覺化是不是主觀的？
- › 想實做自己的視覺化流程，但不知道如何開展，老是做到一半放棄

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難
- › 覺得視覺化需要天生的設計美感？
  - › 好的視覺化是不是主觀的？
- › 想實做自己的視覺化流程，但不知道如何開展，老是做到一半放棄
- › 用 EXCEL 辛苦做出的圖表，卻無法在下次重複使用

# 大部分人對視覺化的常見問題

- › 對於（資料）視覺化想得太簡單或太難
- › 覺得視覺化需要天生的設計美感？
  - › 好的視覺化是不是主觀的？
- › 想實做自己的視覺化流程，但不知道如何開展，老是做到一半放棄
- › 用 EXCEL 辛苦做出的圖表，卻無法在下次重複使用
- › 導入視覺化工具，卻變成「純試試」，一點都不實用，造成導入的軟體無法發揮其效用

目的決定手段的好壞

# WHY DO WE CREATE DATA VIZ?

# PURPOSES OF VIZ

- > COMMUNICATE
- > 報告給老闆、同事.....
- > 讓相關的人做決策
- > ...

# WHAT I OFTEN DO: DATA ANALYSIS (EDA)

# OUTLINES

- > THE LANGUAGE OF DATA VIZ (20 MIN)
- > VIZ THE MODEL BY EXAMPLE (20 MIN)
- > SUMMARY AND RECAP (5 MIN)

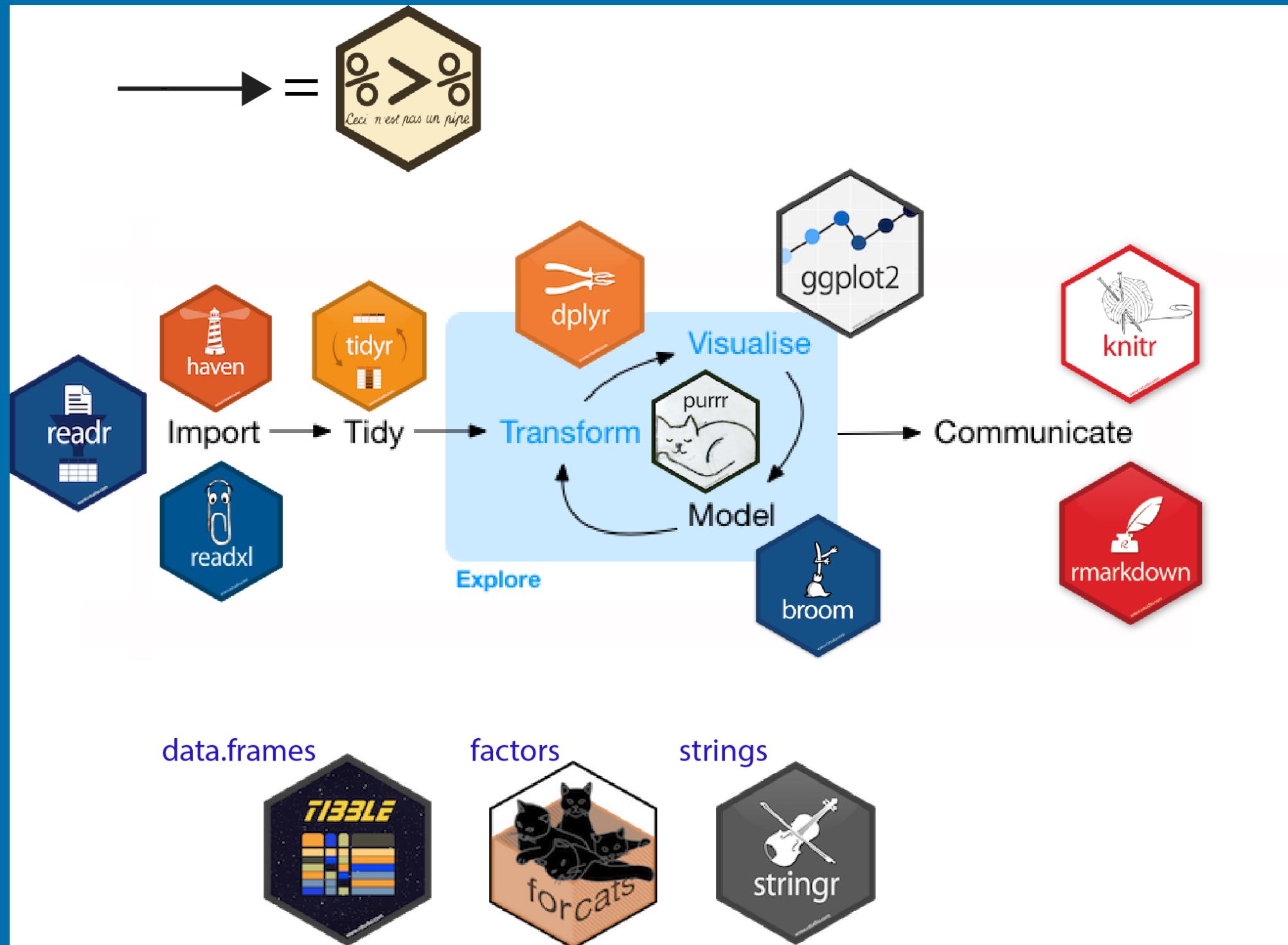
# 今天從心法開始談 THE LANGUAGE OF DATA VIZ

常常我們學到的：

1. 「一張好的圖可以勝過千言萬語」  
› 所以我要怎麼做？
2. XX 圖可以用 YY 方法做  
› 我怎麼知道要畫 XX 圖？

資料視覺化，重點在「資料」

# Tidyverse -- HADLEY WICKHAM



# 資料視覺化的心法

# 圖表如何從無到有？

1. ISSUES: 這張圖表想解決的問題是什麼？

2. GEOMS: 要畫什麼樣的圖？

› BAR CHART, LINE CHART, HEAT MAP, ...

3. VARIABLES: 需要那些變數？

› 我的資料有這些變數嗎？

4. AESTHETIC MAPPING

5. LAYERS

6. STACKING

在 GRAMMAR OF GRAPHICS 裡面最重要的概念就是

AESTHETIC MAPPING

# 最簡單的例子



## BAR CHART

Foreign banks' lending to Russia\*

End-December 2013

\$bn

France

0 10 20 30 40 50

1.7

United States

1.1

Italy

3.5

Austria

4.9

Germany

0.8

Netherlands

1.4

Britain

0.5

Japan

0.5

Sweden

1.0

Spain

0.1

Canada

0.1

India

1.8

Share of total foreign lending, %

\*Adjusted for credit risk

Source: BIS

# 最簡單的例子



Foreign banks' lending to Russia\*

End-December 2013

\$bn

France

0 10 20 30 40 50

Share of total foreign lending, %

1.7

United States

1.1

Italy

3.5

Austria

4.9

Germany

0.8

Netherlands

1.4

Britain

0.5

Japan

0.5

Sweden

1.0

Spain

0.1

Canada

0.1

India

1.8

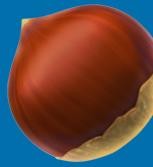
Source: BIS

\*Adjusted for credit risk

BAR CHART

> GEOM: BAR

# 最簡單的例子



## Foreign banks' lending to Russia\*

End-December 2013

\$bn

France

United States

Italy

Austria

Germany

Netherlands

Britain

Japan

Sweden

Spain

Canada

India

Share of total foreign lending, %

1.7

1.1

3.5

4.9

0.8

1.4

0.5

0.5

1.0

0.1

0.1

1.8

Source: BIS

\*Adjusted for credit risk

## BAR CHART

- > GEOM: BAR
- > X: PERCENTAGE

# 最簡單的例子



Foreign banks' lending to Russia\*

End-December 2013

\$bn

France

0 10 20 30 40 50

Share of total foreign lending, %

1.7

United States

1.1

Italy

3.5

Austria

4.9

Germany

0.8

Netherlands

1.4

Britain

0.5

Japan

0.5

Sweden

1.0

Spain

0.1

Canada

0.1

India

1.8

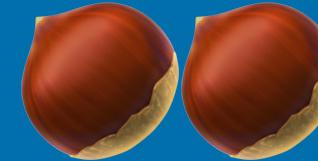
Source: BIS

\*Adjusted for credit risk

## BAR CHART

- > GEOM: BAR
- > X: PERCENTAGE
- > Y: FOREIGN COUNTRIES

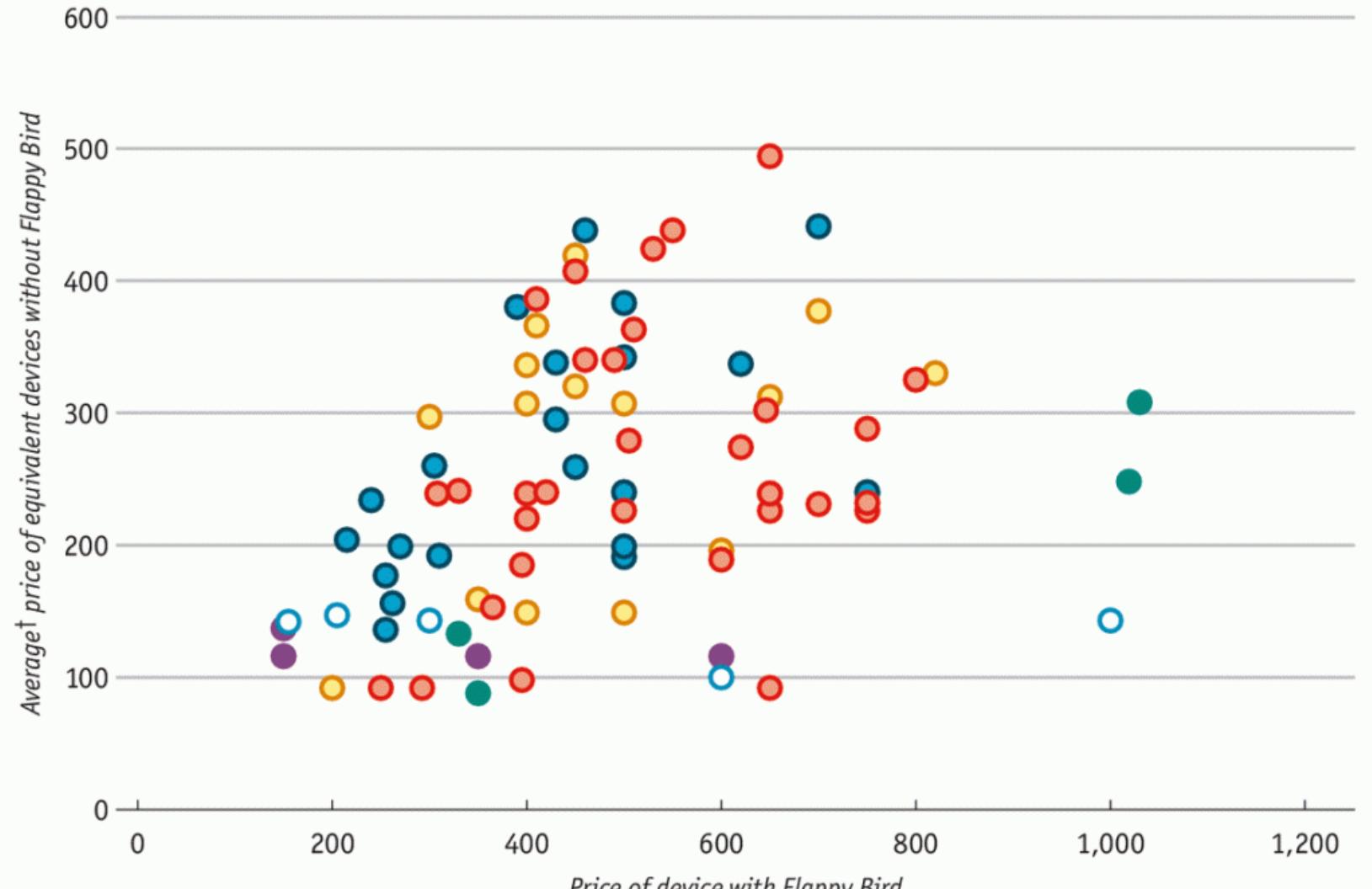
# 例子1



## Estimating the value of Flappy Bird

Prices on Ebay\*, February 14th 2014, £

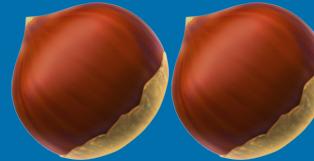
Ipad Iphone 5 Iphone 4 Ipod  
Samsung Galaxy S4 Samsung Galaxy S3



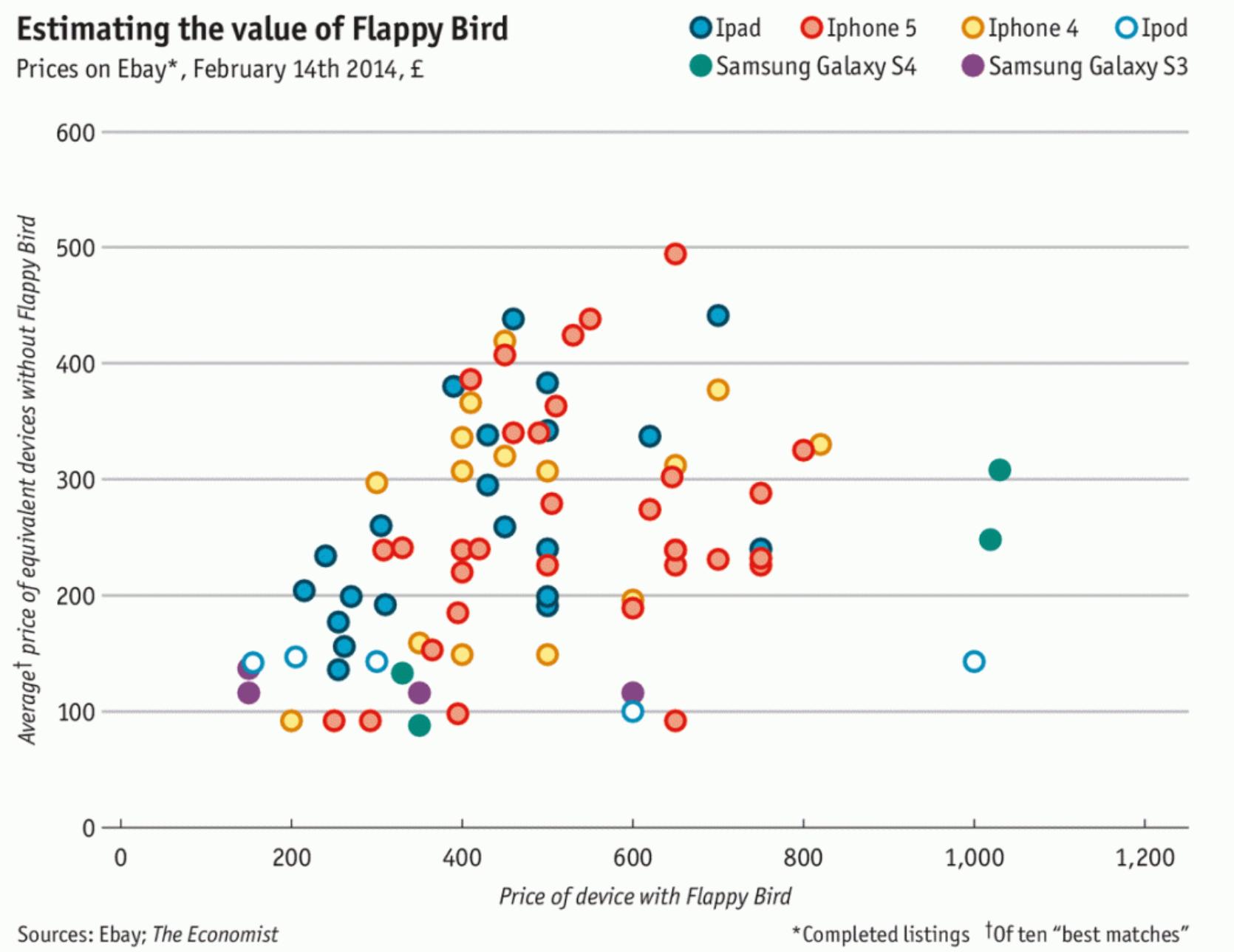
Sources: Ebay; The Economist

\*Completed listings †Of ten "best matches"

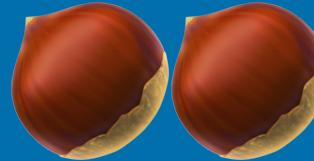
# 例子1



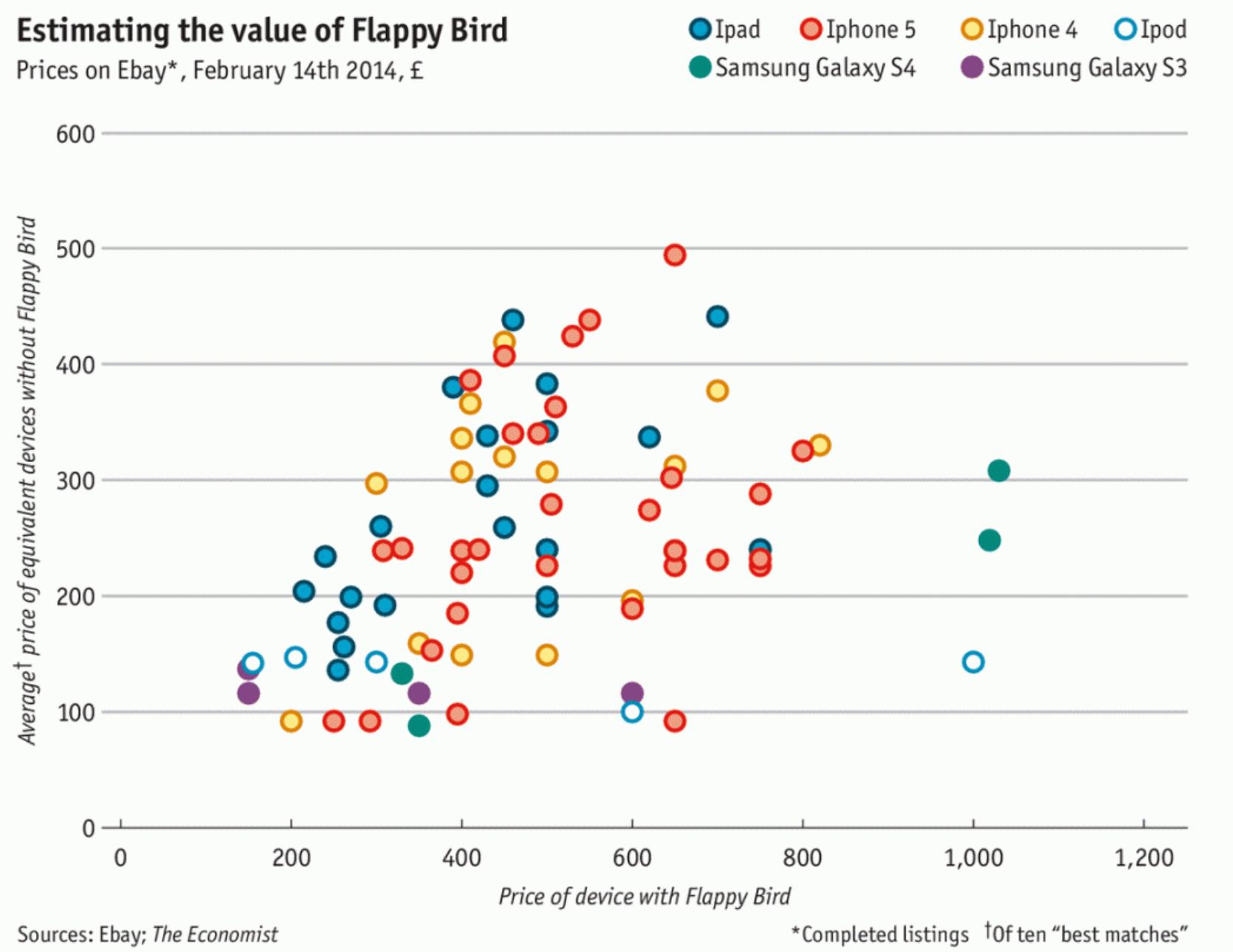
> GEOM: POINT



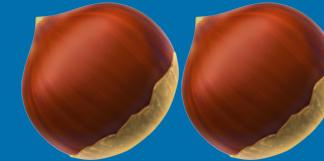
# 例子1



- > GEOM: POINT
- > X: PRICE



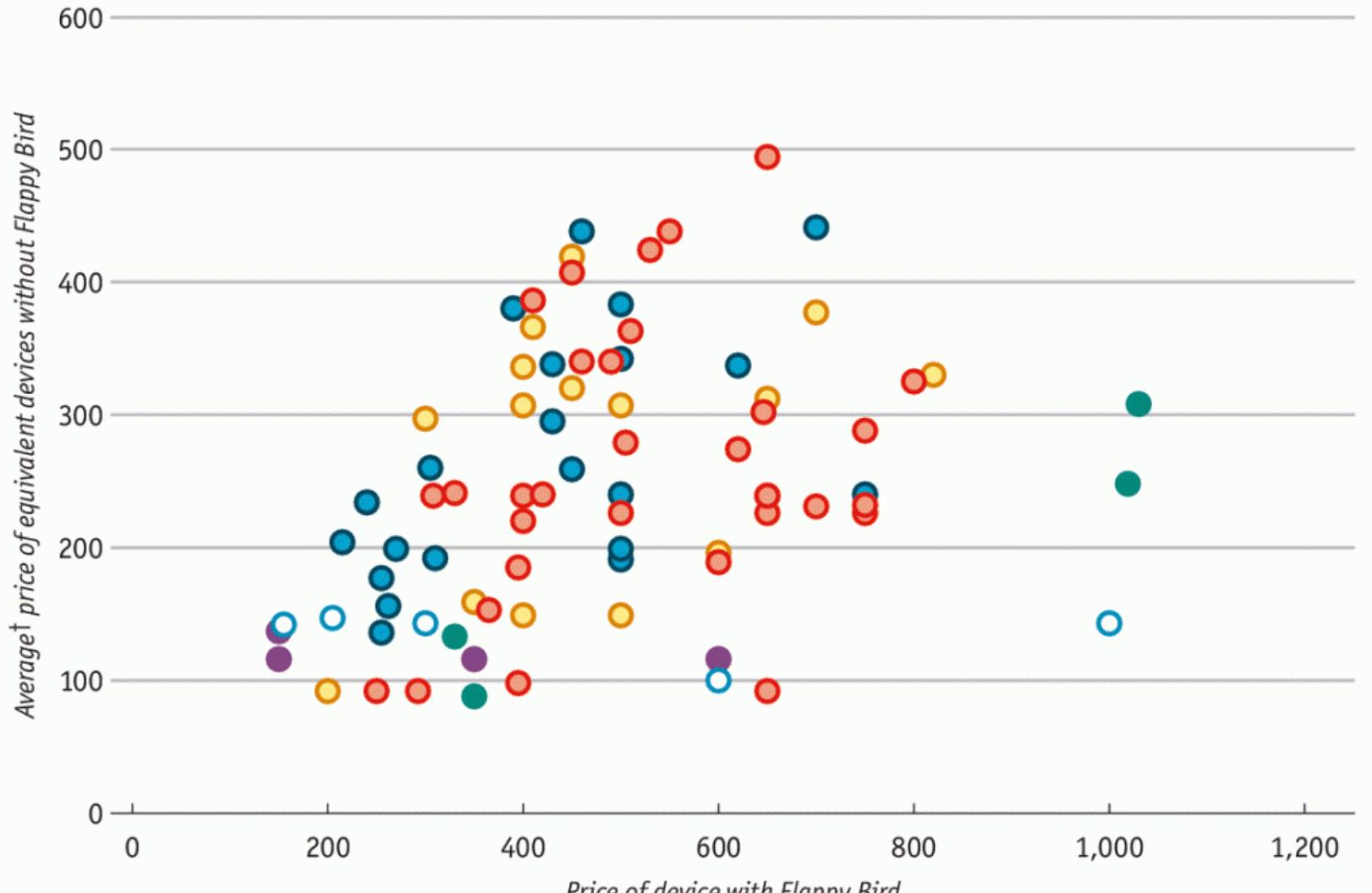
# 例子1



## Estimating the value of Flappy Bird

Prices on Ebay\*, February 14th 2014, £

Ipad Iphone 5 iPhone 4 Ipod  
Samsung Galaxy S4 Samsung Galaxy S3

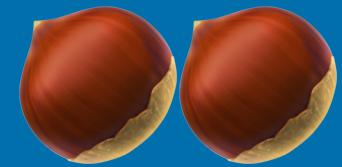


> GEOM: POINT

> X: PRICE

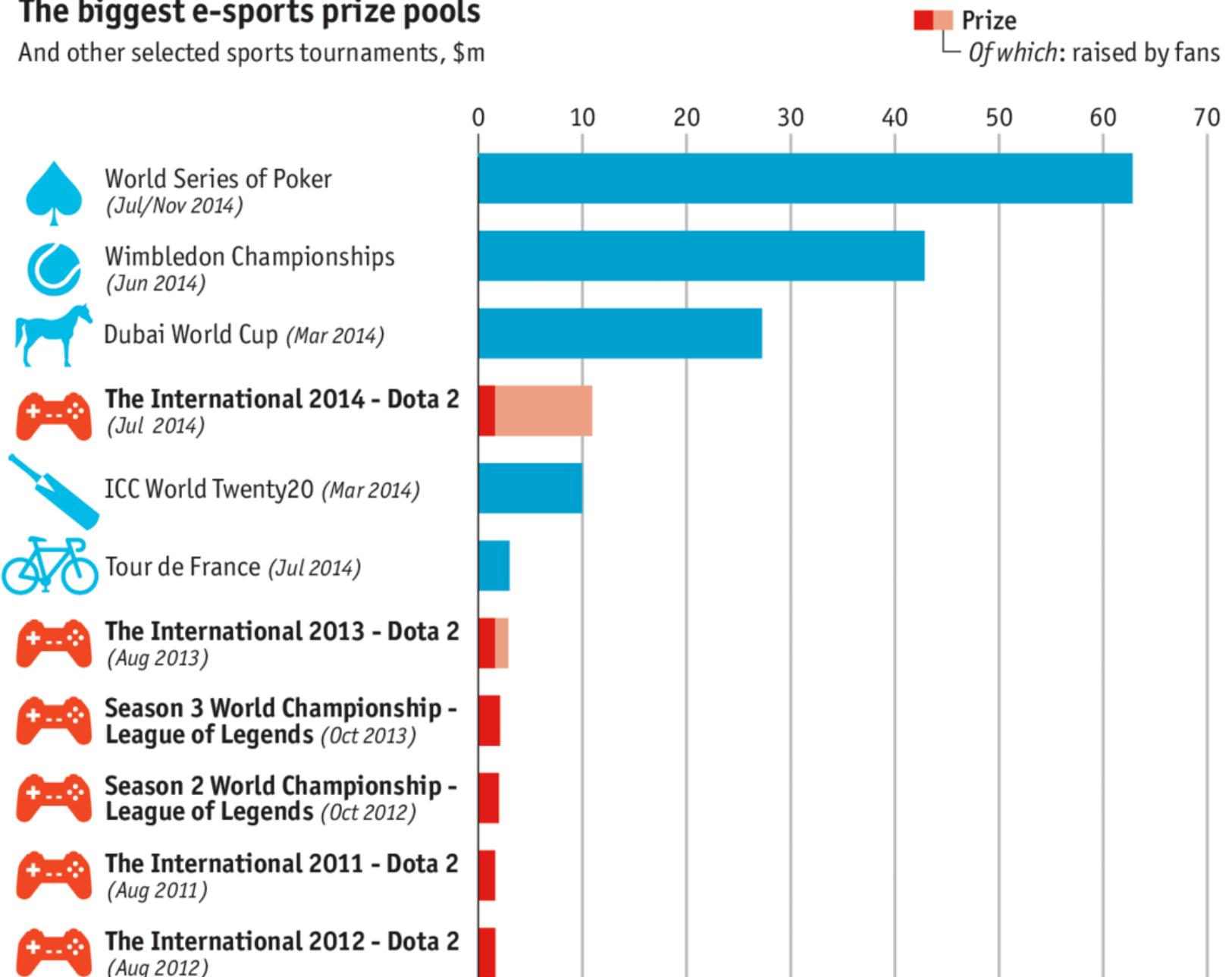
> Y: AVG PRICE WITHOUT FLAPPY BIRD

# 例子2



## The biggest e-sports prize pools

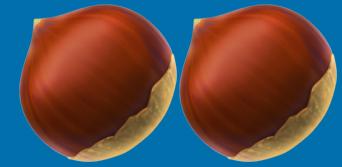
And other selected sports tournaments, \$m



Sources: E-Sports Earnings; company reports; press reports

Economist.com/graphicdetail

# 例子2



## > GEOM: BAR

### The biggest e-sports prize pools

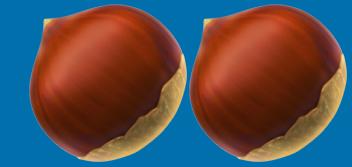
And other selected sports tournaments, \$m



Sources: E-Sports Earnings; company reports; press reports

Economist.com/graphicdetail

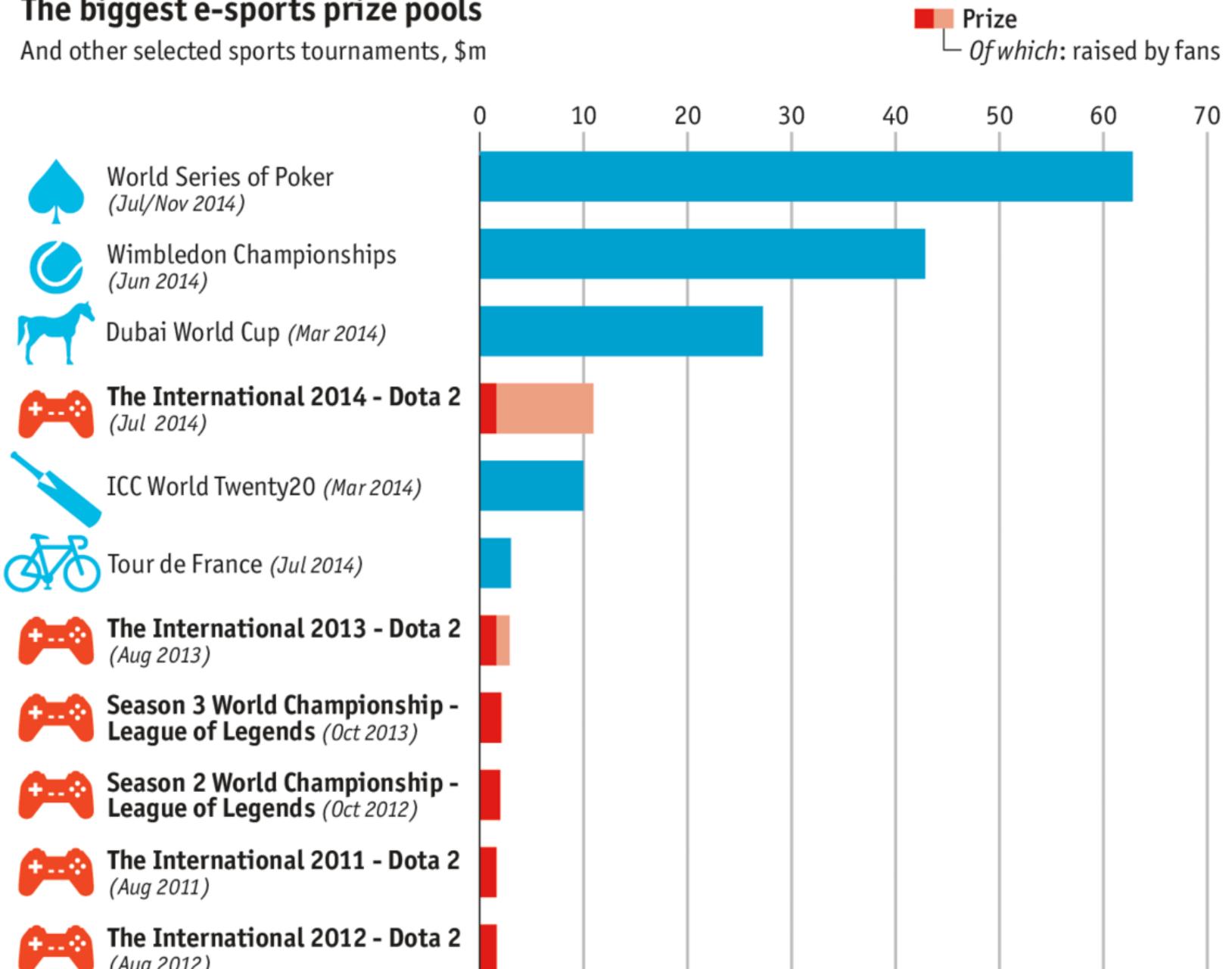
# 例子2



- > GEOM: BAR
- > X: PRIZE POOL

## The biggest e-sports prize pools

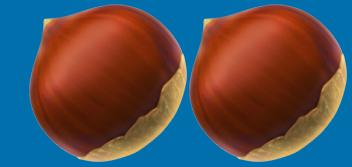
And other selected sports tournaments, \$m



Sources: E-Sports Earnings; company reports; press reports

Economist.com/graphicdetail

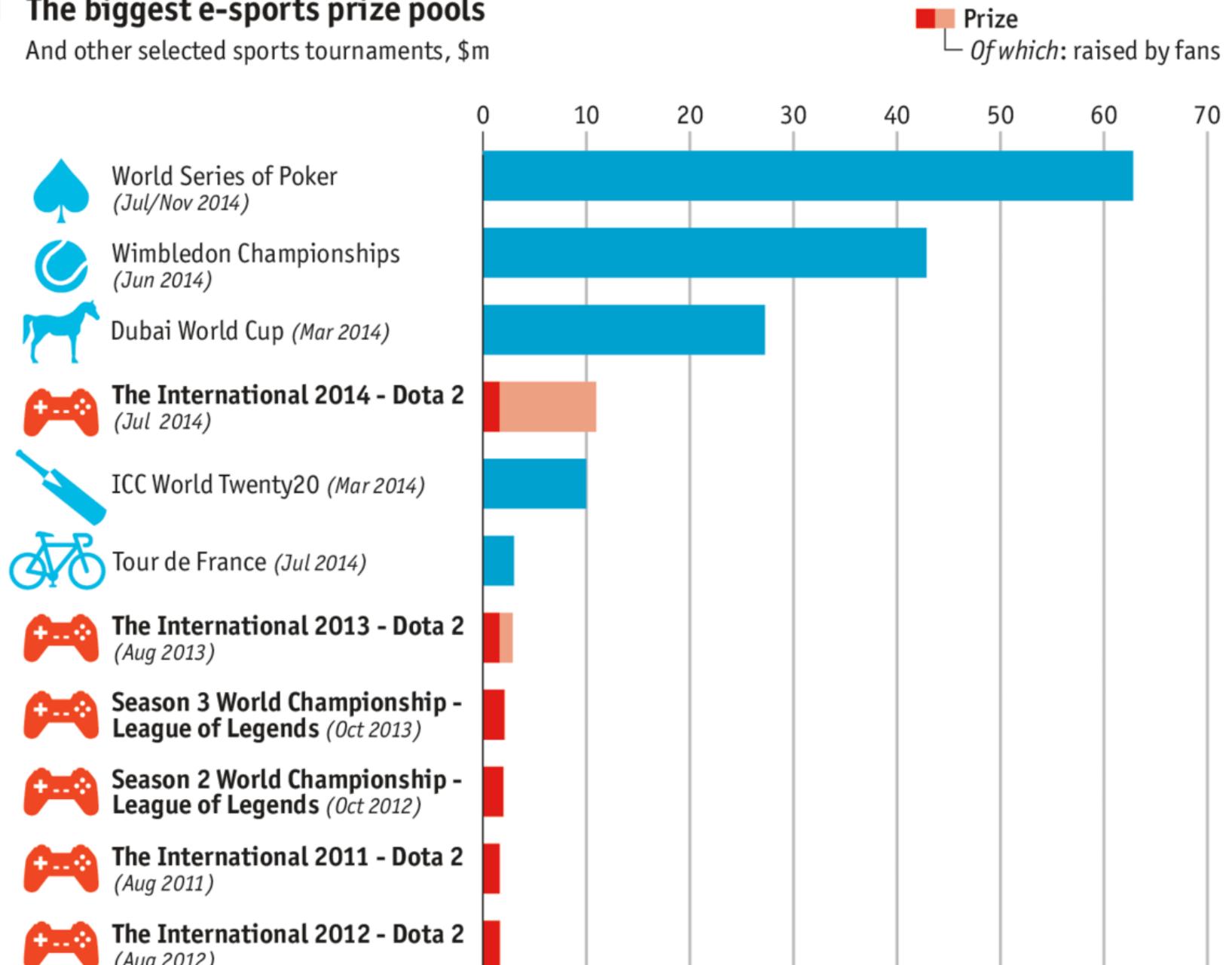
# 例子2



- > GEOM: BAR
- > X: PRIZE POOL
- > Y: E-SPORTS

## The biggest e-sports prize pools

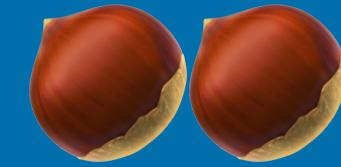
And other selected sports tournaments, \$m



Sources: E-Sports Earnings; company reports; press reports

Economist.com/graphicdetail

# 例子2

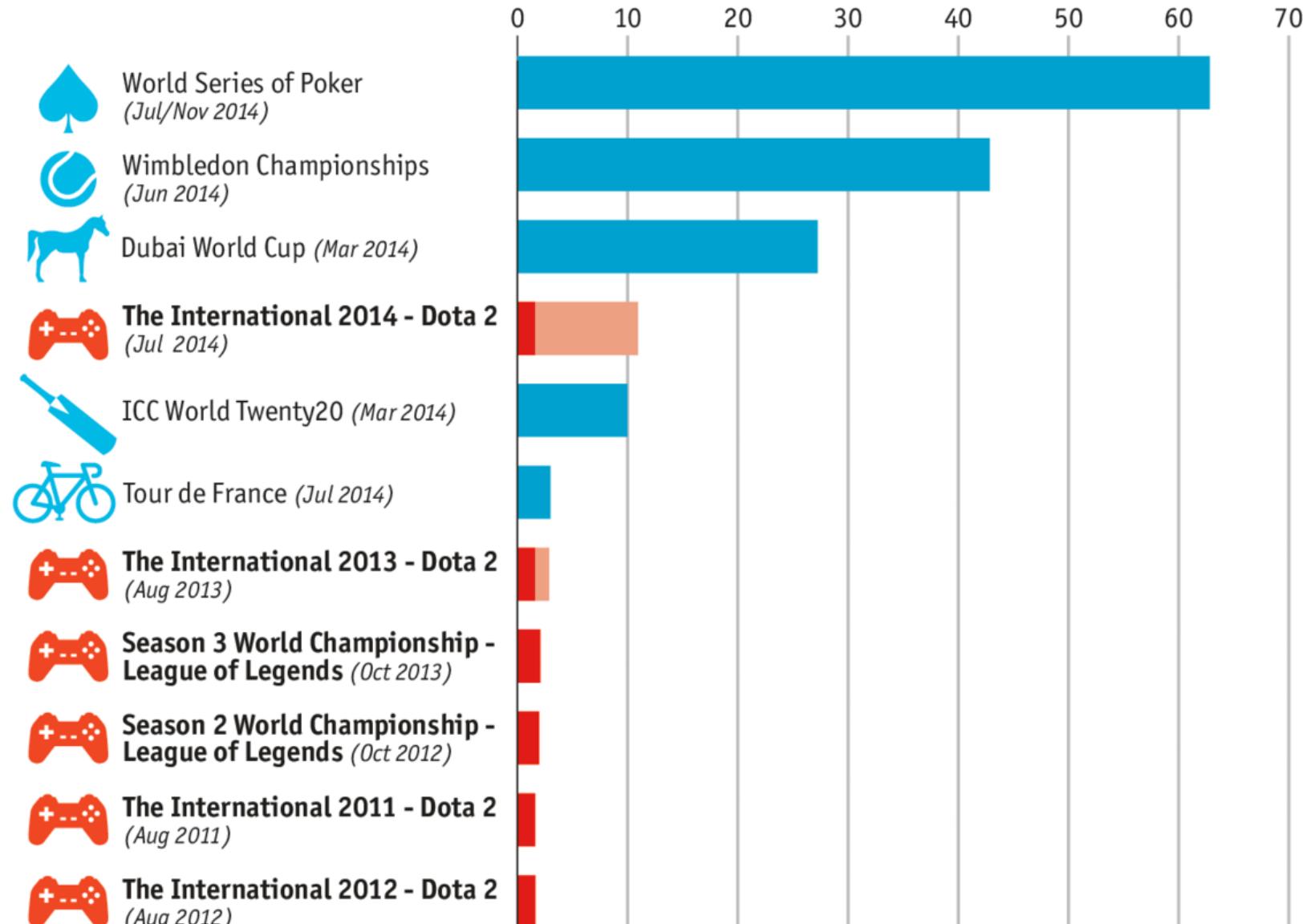


- > GEOM: BAR
- > X: PRIZE POOL
- > Y: E-SPORTS
- > FILL COLOR: RAISED BY FAN / NOT FAN

## The biggest e-sports prize pools

And other selected sports tournaments, \$m

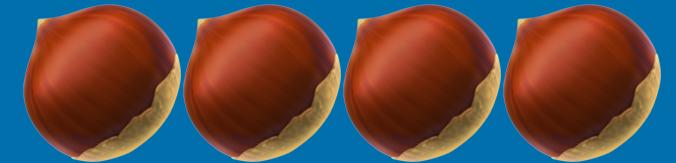
Prize  
Of which: raised by fans



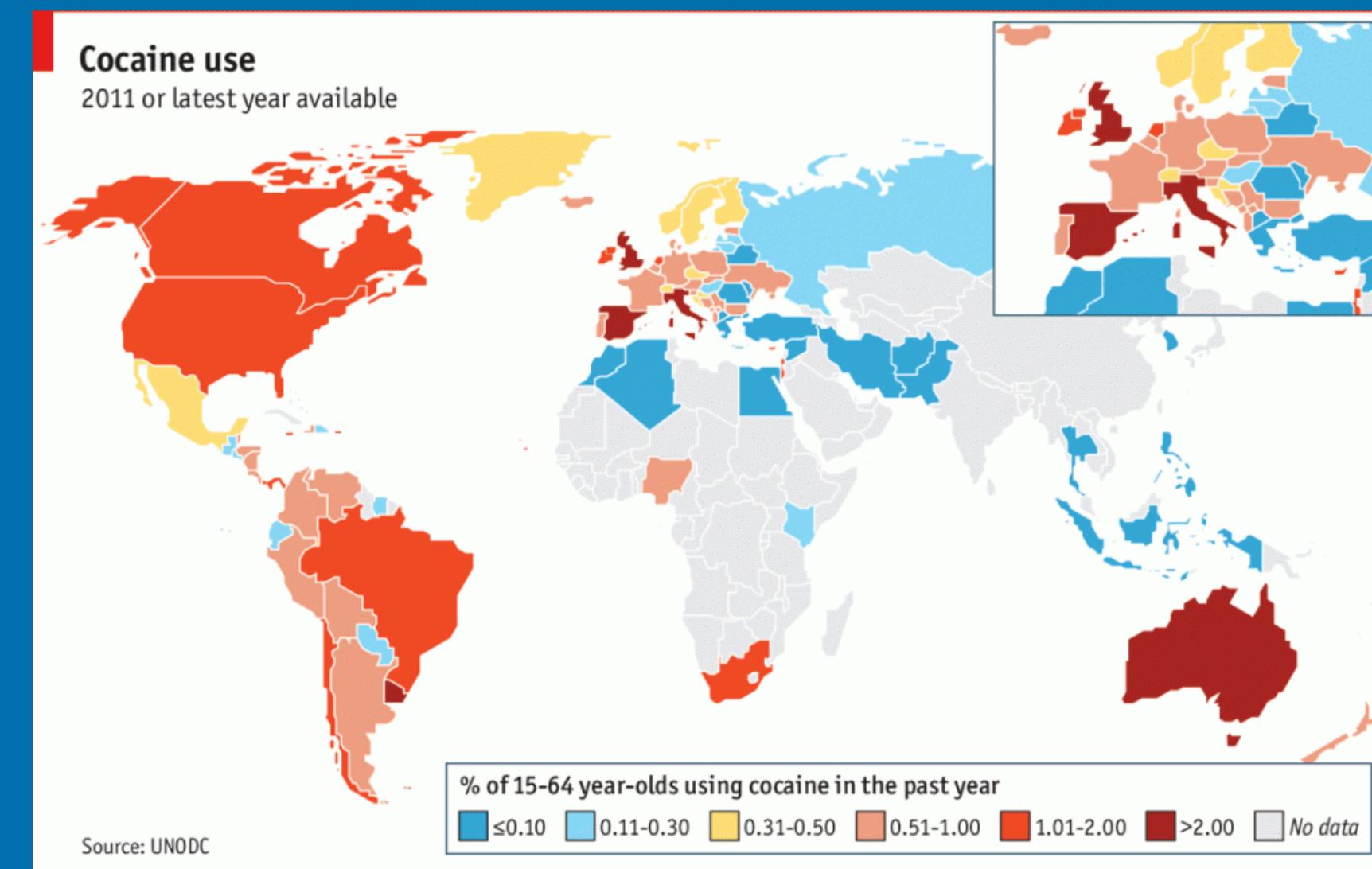
Sources: E-Sports Earnings; company reports; press reports

Economist.com/graphicdetail

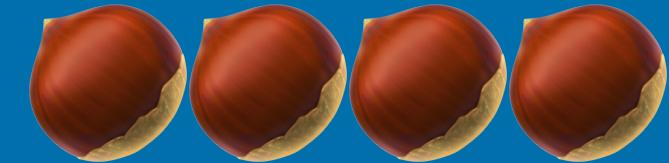
# 例子3



地圖

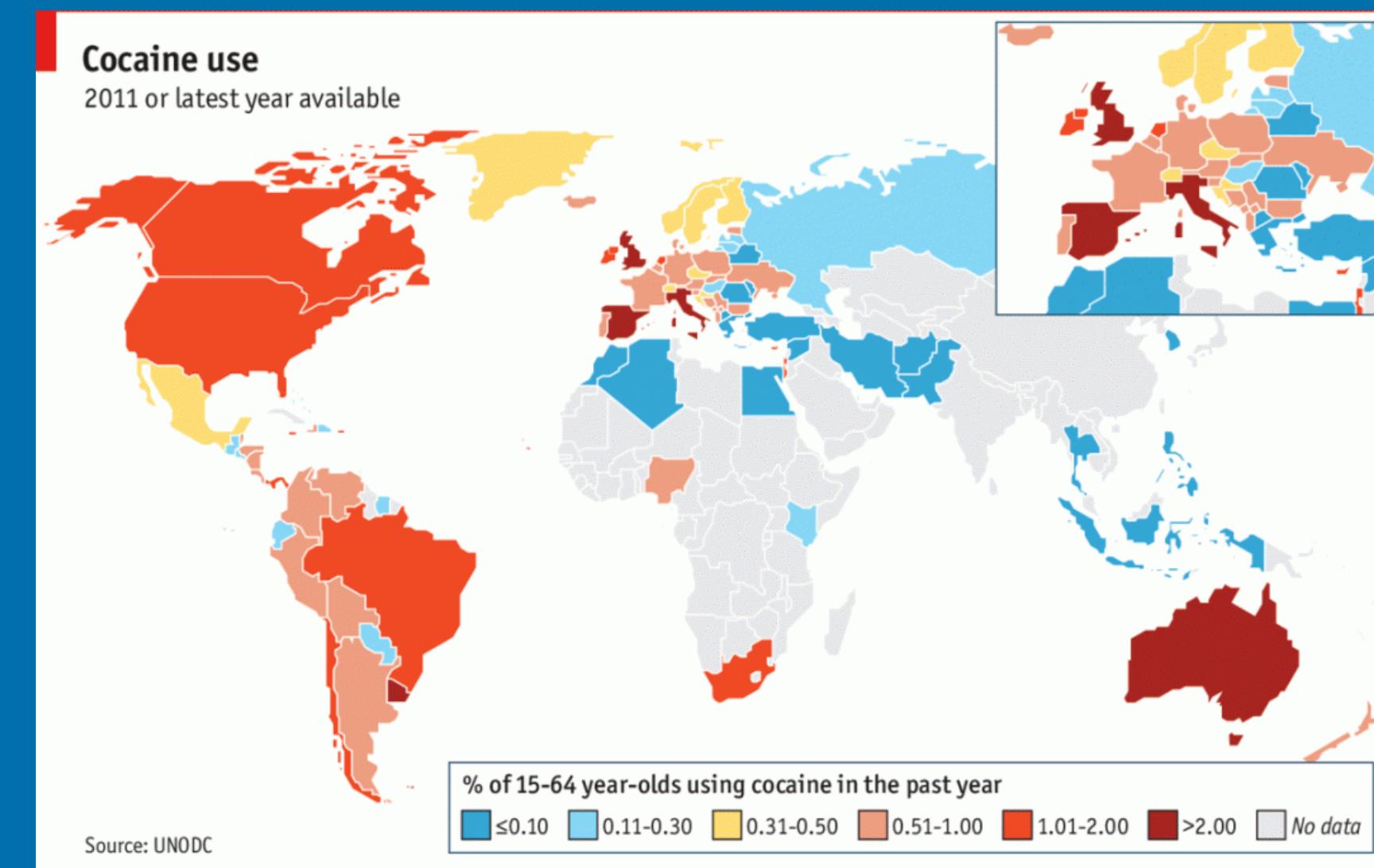


# 例子3

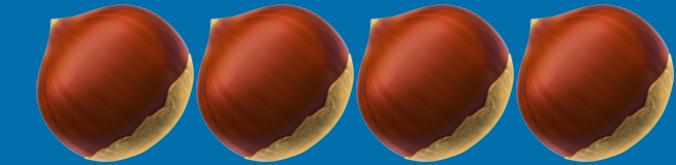


地圖

> GEOM: MAP RASTER

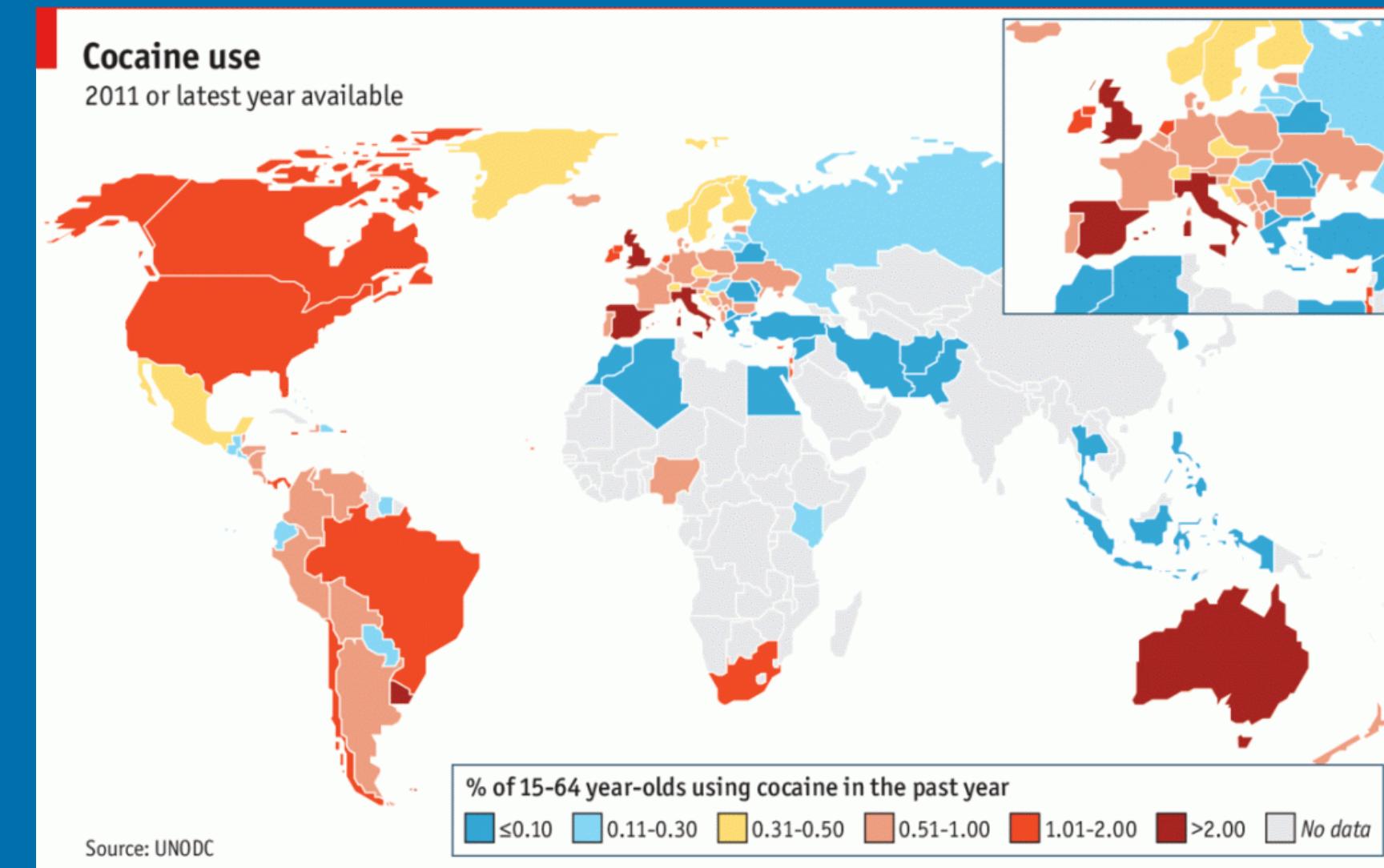


# 例子3

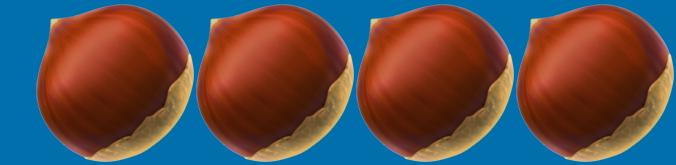


## 地圖

- › GEOM: MAP RASTER
- › RASTER: COUNTRIES

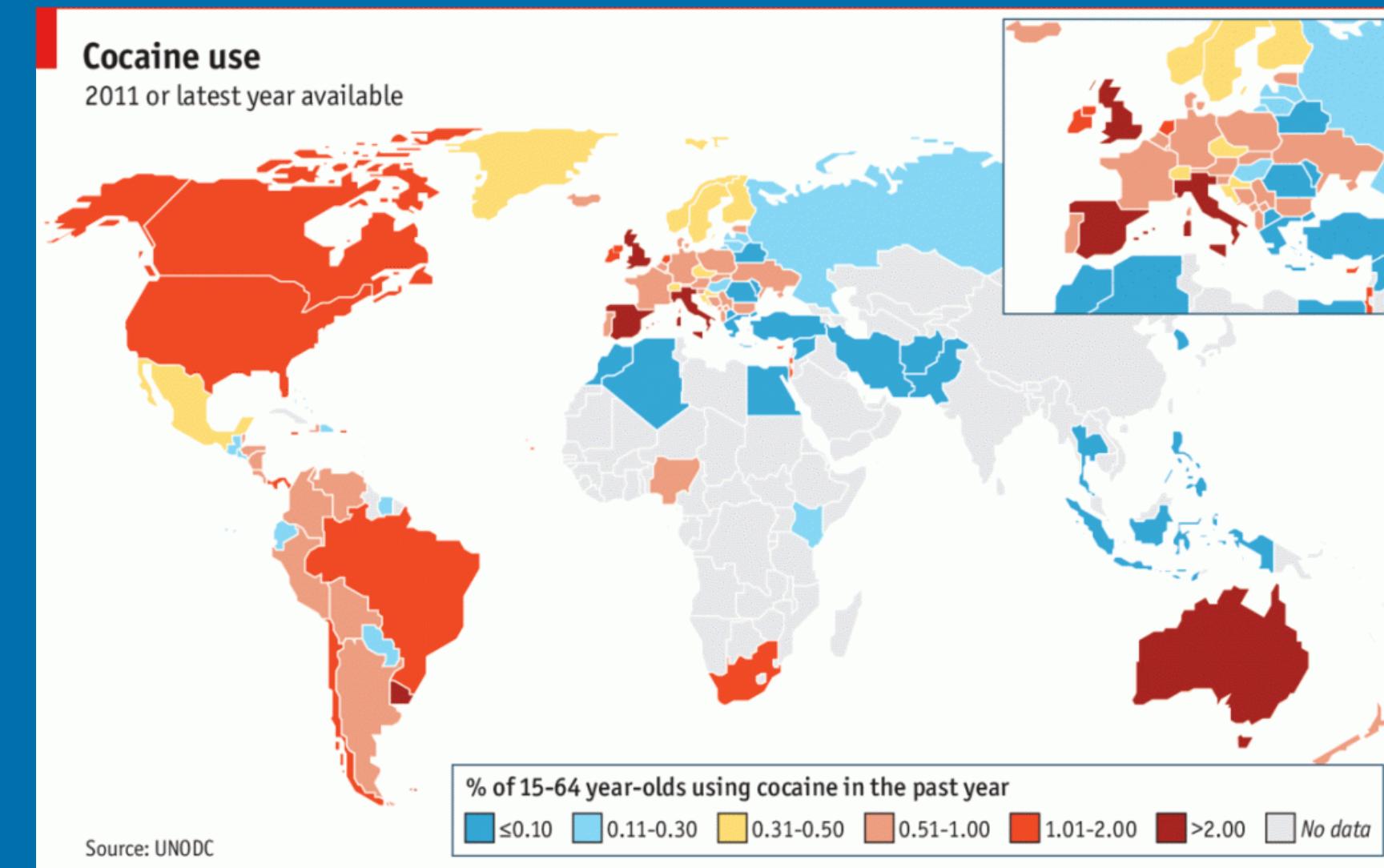


# 例子3



## 地圖

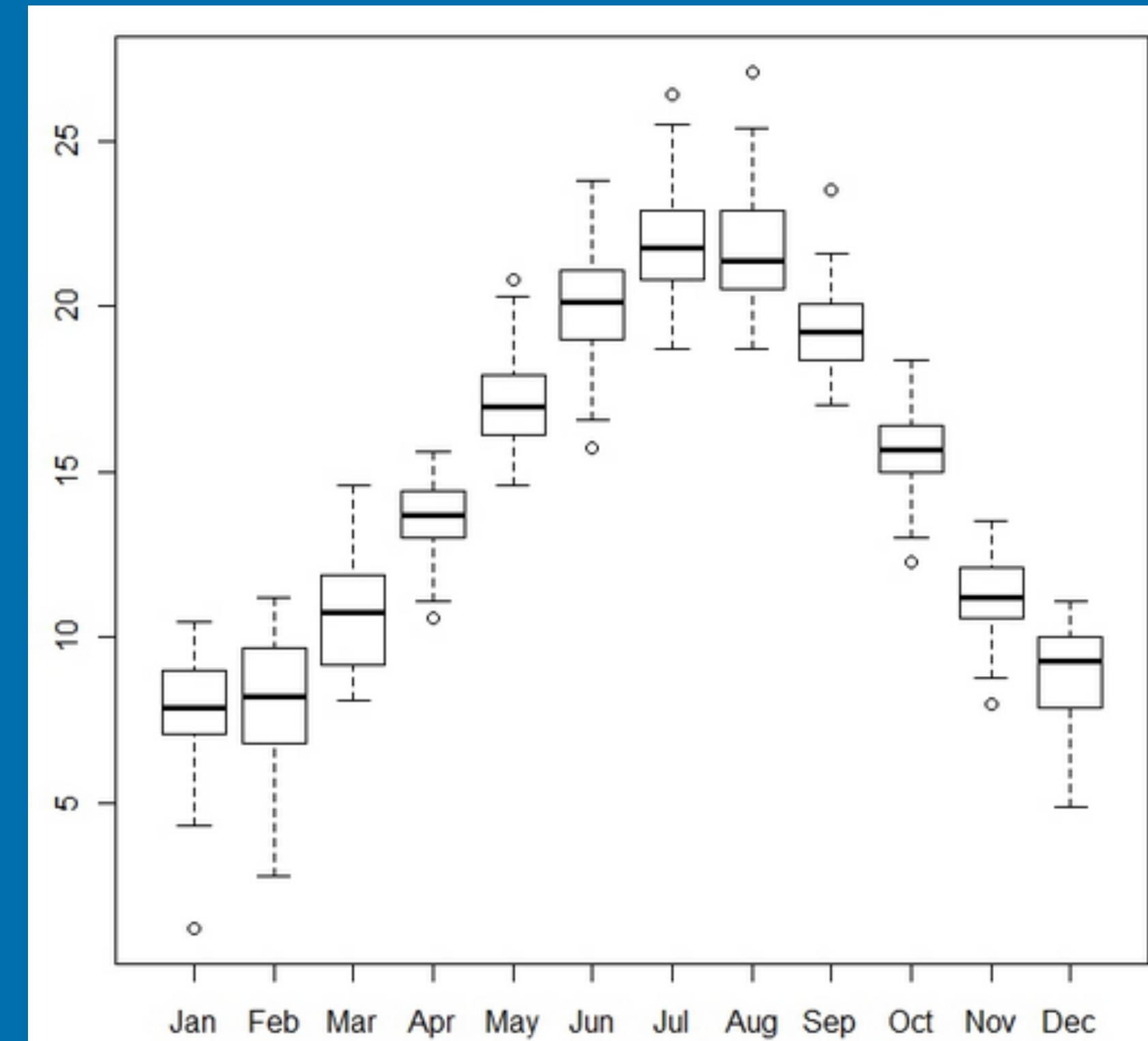
- > GEOM: MAP RASTER
- > RASTER: COUNTRIES
- > FILL COLOR: PERCENTAGE



# 例子4



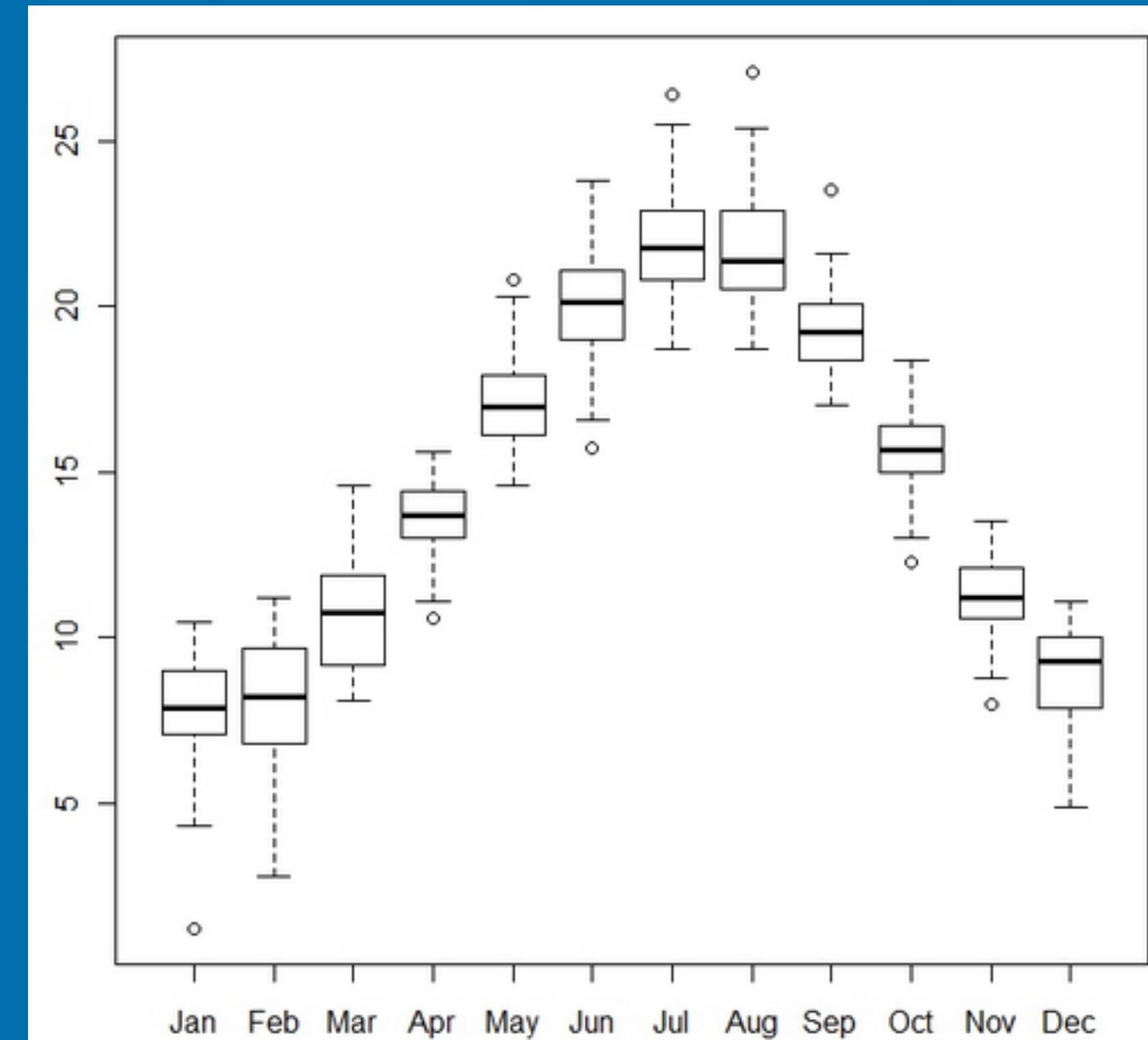
BOXPLOT -- JOHN W. TUKEY



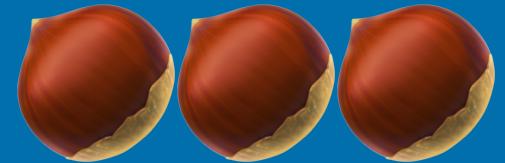
# 例子4



BOXPLOT -- JOHN W. TUKEY  
➤ STATISTICAL PURPOSE

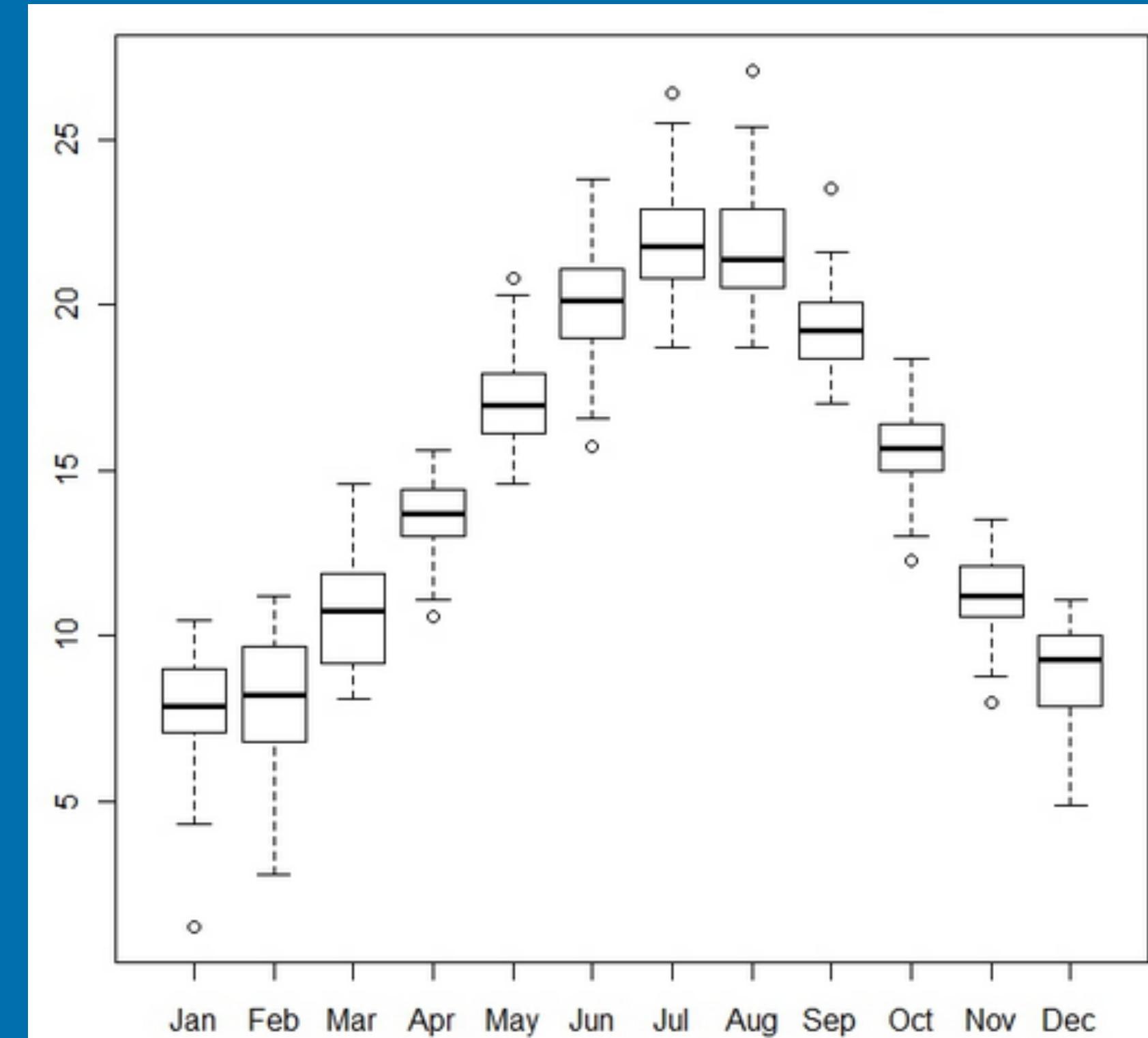


# 例子4

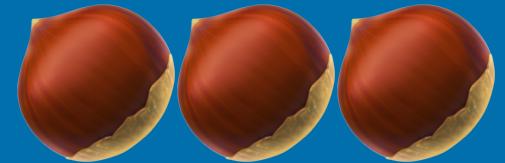


BOXPLOT -- JOHN W. TUKEY

- › STATISTICAL PURPOSE
- › GEOM: BOXPLOT

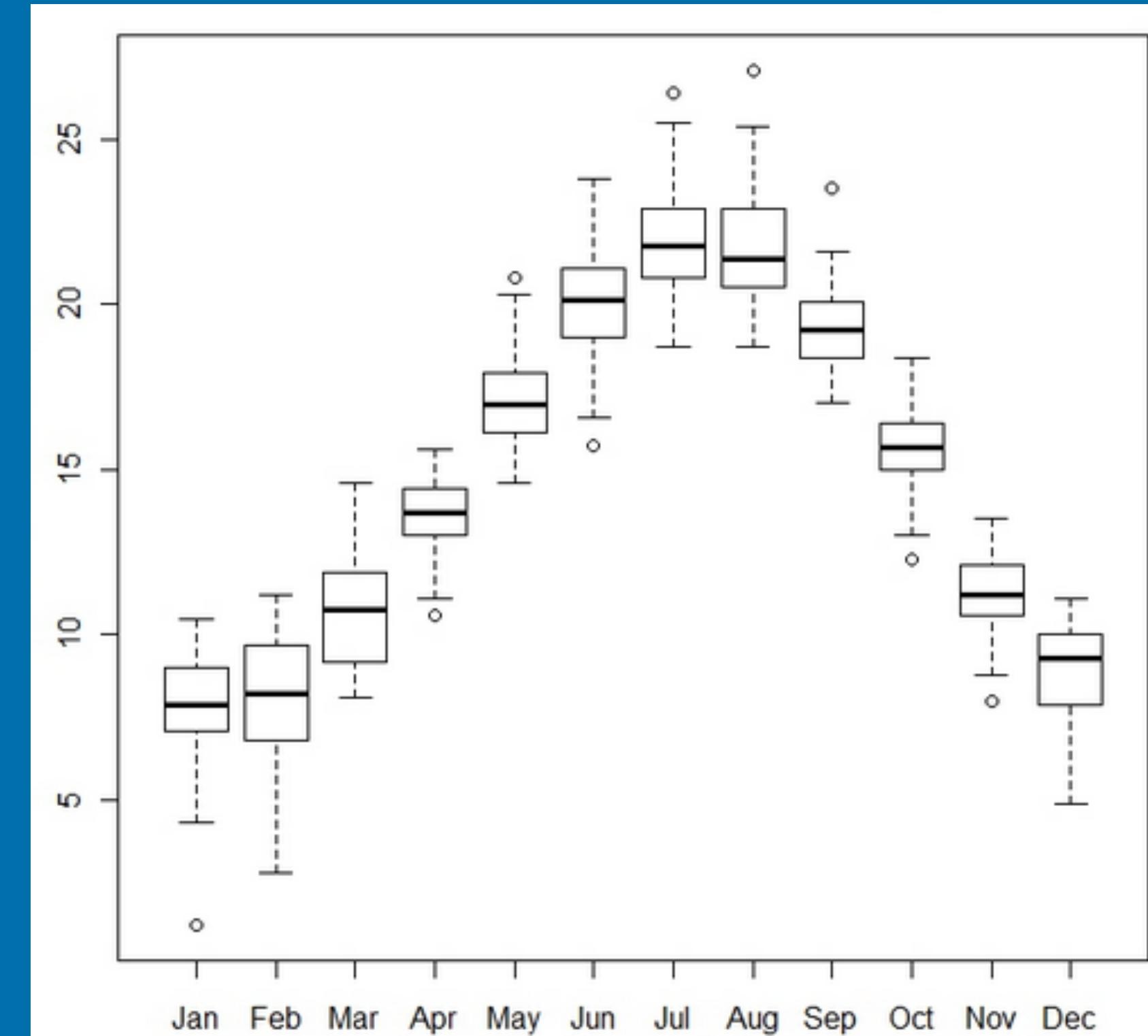


# 例子4

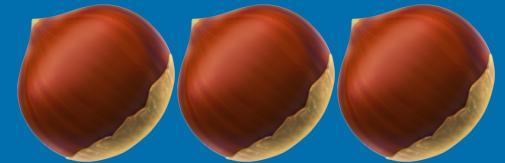


BOXPLOT -- JOHN W. TUKEY

- > STATISTICAL PURPOSE
- > GEOM: BOXPLOT
- > X: ...

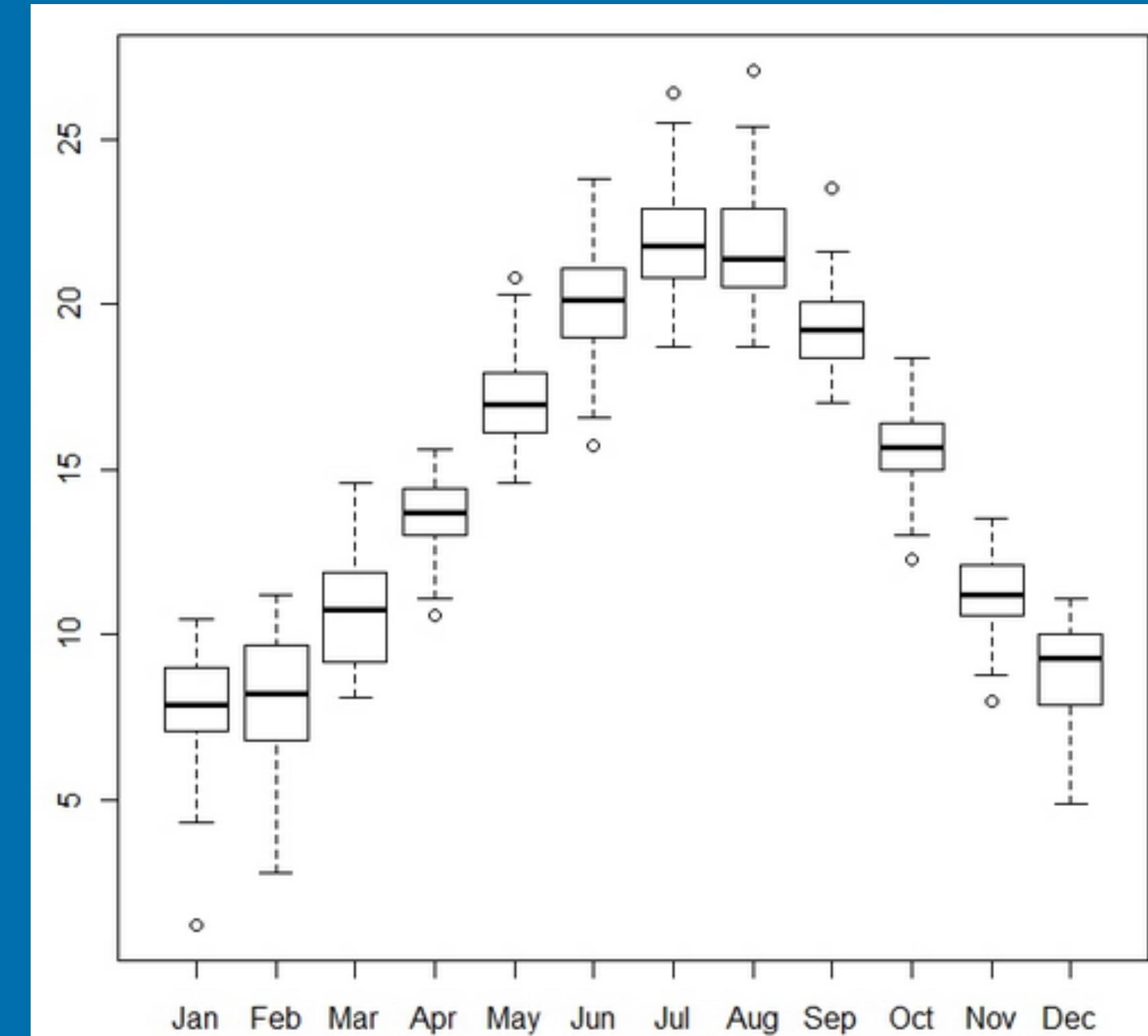


# 例子4

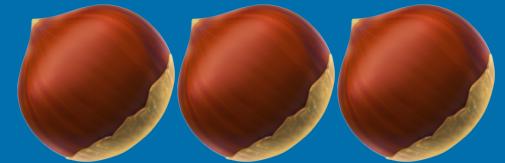


BOXPLOT -- JOHN W. TUKEY

- > STATISTICAL PURPOSE
- > GEOM: BOXPLOT
- > X: ...
- > Y: ...

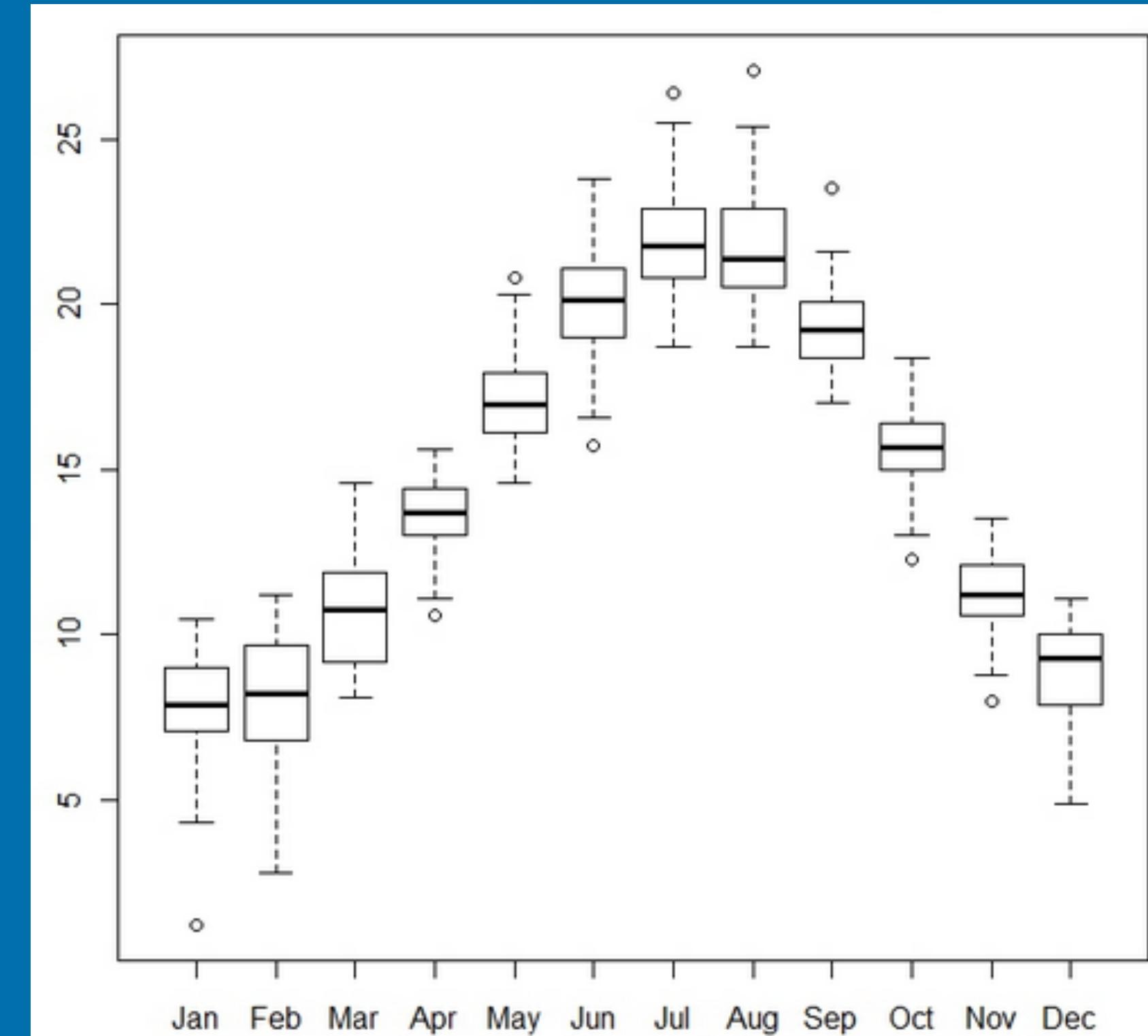


# 例子4

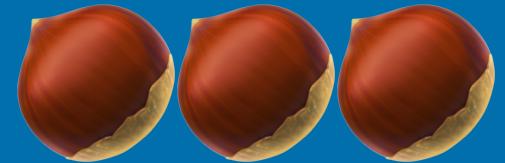


BOXPLOT -- JOHN W. TUKEY

- > STATISTICAL PURPOSE
- > GEOM: BOXPLOT
  - > X: ...
  - > Y: ...
- > 1ST QUANTILE

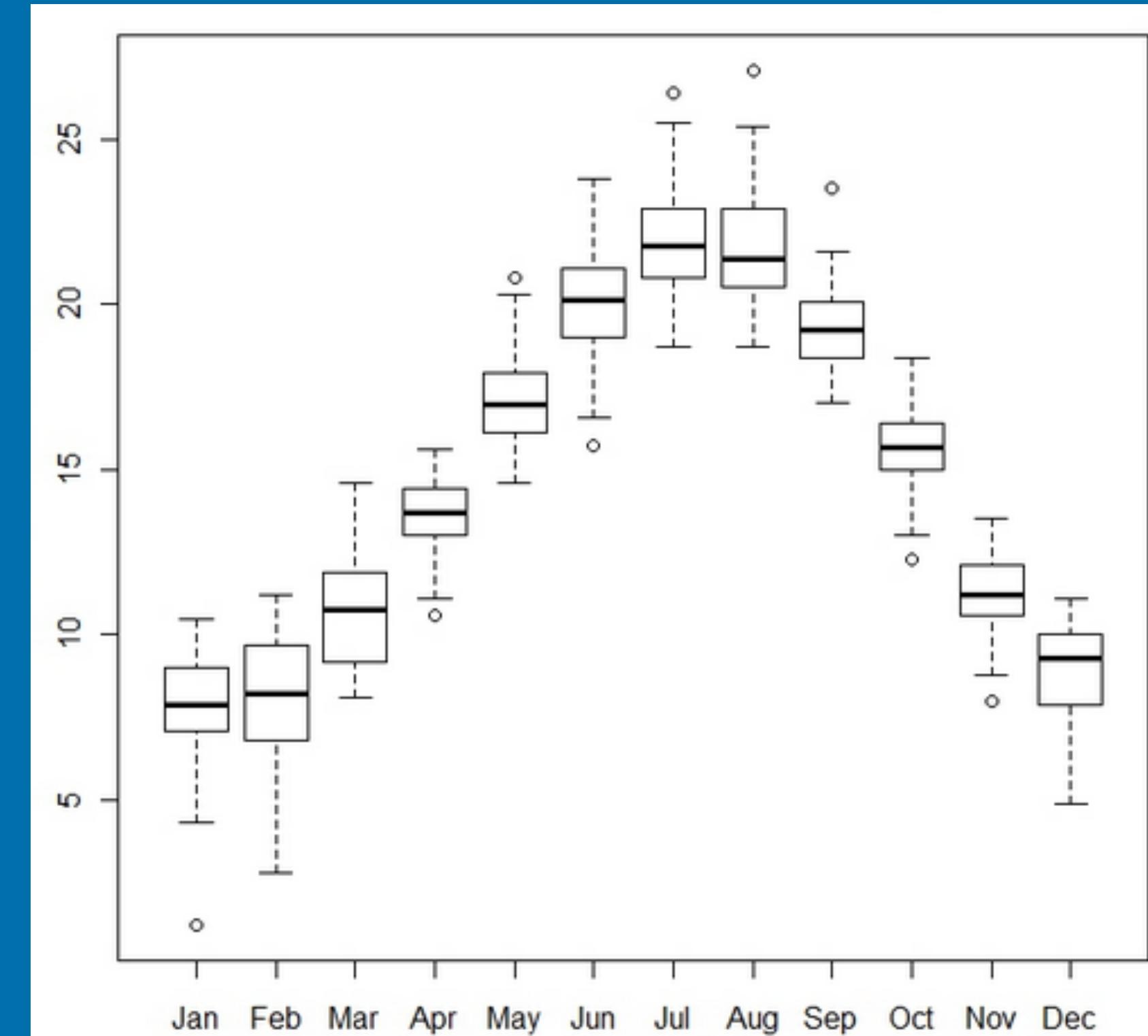


# 例子4

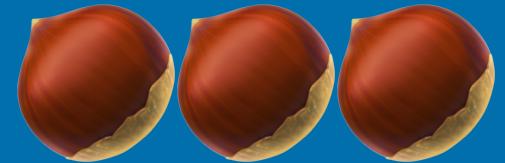


BOXPLOT -- JOHN W. TUKEY

- > STATISTICAL PURPOSE
- > GEOM: BOXPLOT
  - > X: ...
  - > Y: ...
- > 1ST QUANTILE
- > MEDIAN

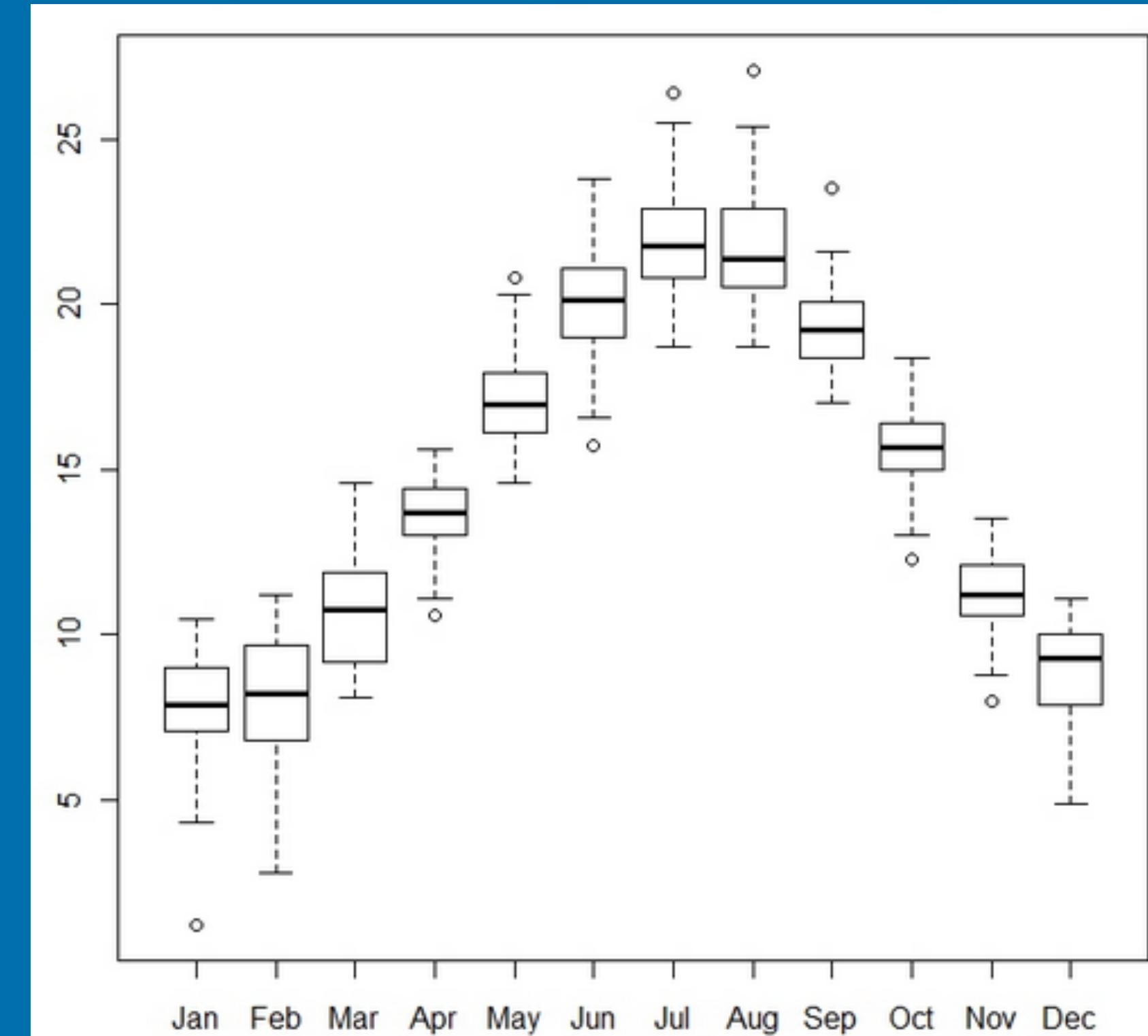


# 例子4

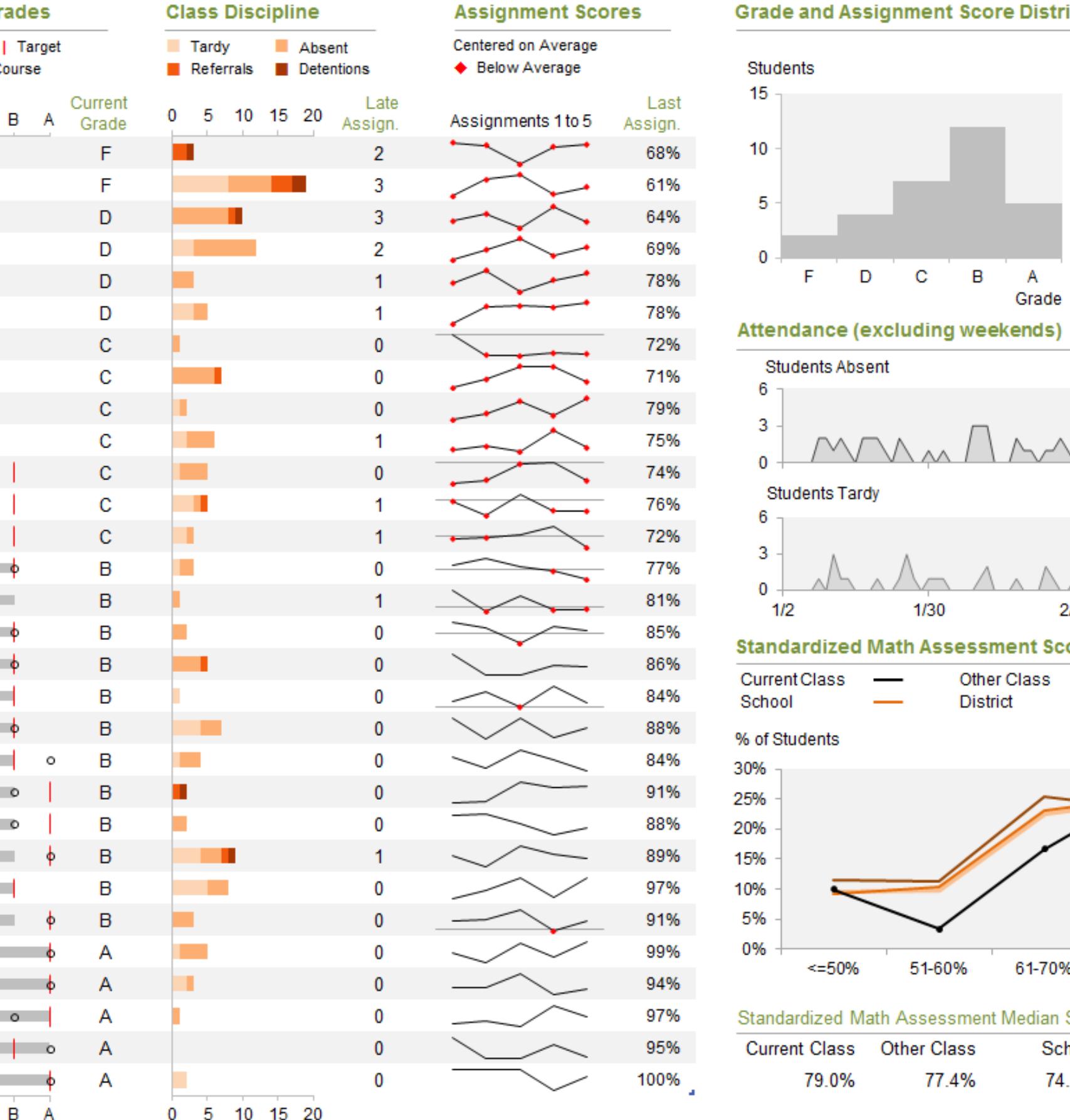


BOXPLOT -- JOHN W. TUKEY

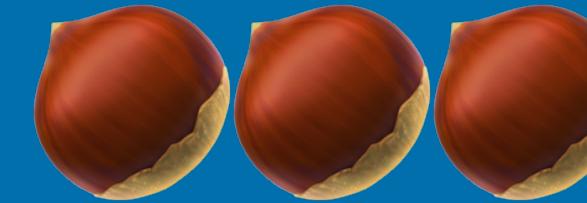
- > STATISTICAL PURPOSE
- > GEOM: BOXPLOT
  - > X: ...
  - > Y: ...
- > 1ST QUANTILE
- > MEDIAN
- > 3RD QUANTILE



## Grade 10 Algebra Course



# 例子5



## STEPHEN FEW -- DASHBOARD

> GEOM: TOO MANY

> GRID

> ...

# 例子6



表格??

## Best cities

Based on liveability index\*

Rank	City	Country
1	Melbourne	Australia
2	Vienna	Austria
3	Vancouver	Canada
4	Toronto	Canada
5	Calgary	Canada
5	Adelaide	Australia
7	Sydney	Australia
8	Helsinki	Finland
9	Perth	Australia
10	Auckland	New Zealand

\* Based on 30 factors spread across five areas: stability, infrastructure, education, health care and environment

Source: Economist Intelligence Unit

# 例子6



表格??

› GEOM: 表格

## Best cities

Based on liveability index\*

Rank	City	Country
1	Melbourne	Australia
2	Vienna	Austria
3	Vancouver	Canada
4	Toronto	Canada
5	Calgary	Canada
5	Adelaide	Australia
7	Sydney	Australia
8	Helsinki	Finland
9	Perth	Australia
10	Auckland	New Zealand

\* Based on 30 factors spread across five areas: stability, infrastructure, education, health care and environment

Source: Economist Intelligence Unit

# 例子6



表格??

- > GEOM: 表格
- > X: VARIABLES

## Best cities

Based on liveability index\*

Rank	City	Country
1	Melbourne	Australia
2	Vienna	Austria
3	Vancouver	Canada
4	Toronto	Canada
5	Calgary	Canada
5	Adelaide	Australia
7	Sydney	Australia
8	Helsinki	Finland
9	Perth	Australia
10	Auckland	New Zealand

\* Based on 30 factors spread across five areas: stability, infrastructure, education, health care and environment

Source: Economist Intelligence Unit

# 例子6



表格??

- › GEOM: 表格
- › X: VARIABLES
- › Y: RECORDS

## Best cities

Based on liveability index\*

Rank	City	Country
1	Melbourne	Australia
2	Vienna	Austria
3	Vancouver	Canada
4	Toronto	Canada
5	Calgary	Canada
5	Adelaide	Australia
7	Sydney	Australia
8	Helsinki	Finland
9	Perth	Australia
10	Auckland	New Zealand

\* Based on 30 factors spread across five areas: stability, infrastructure, education, health care and environment

Source: Economist Intelligence Unit

# 視覺化心法的語言實現 R 的 ggplot2 實做

# GGPLOT2 IN R

- › GGPLOT2 是一個很強大的資料探索及視覺化工具，是最有影響力的 R 套件開發者 HADLEY WICKHAM 所開發
- › 所有繪圖函數都有背後的視覺化邏輯 (GRAMMAR OF GRAPHICS)

**GRAMMAR OF GRAPHICS** 的作用就是幫助我們將圖表拆解成個別元素，然後將這些元素按照邏輯個別操作，正確又簡單地達到圖表的目的

A wide-angle photograph of a mountainous landscape. In the foreground, there are dark, rocky slopes covered with sparse, dry vegetation. A prominent, rounded peak rises in the center, its slopes covered in dense green forests. The background features more mountain ridges, fading into a hazy blue under a clear, light blue sky.

# LIVE DEMO

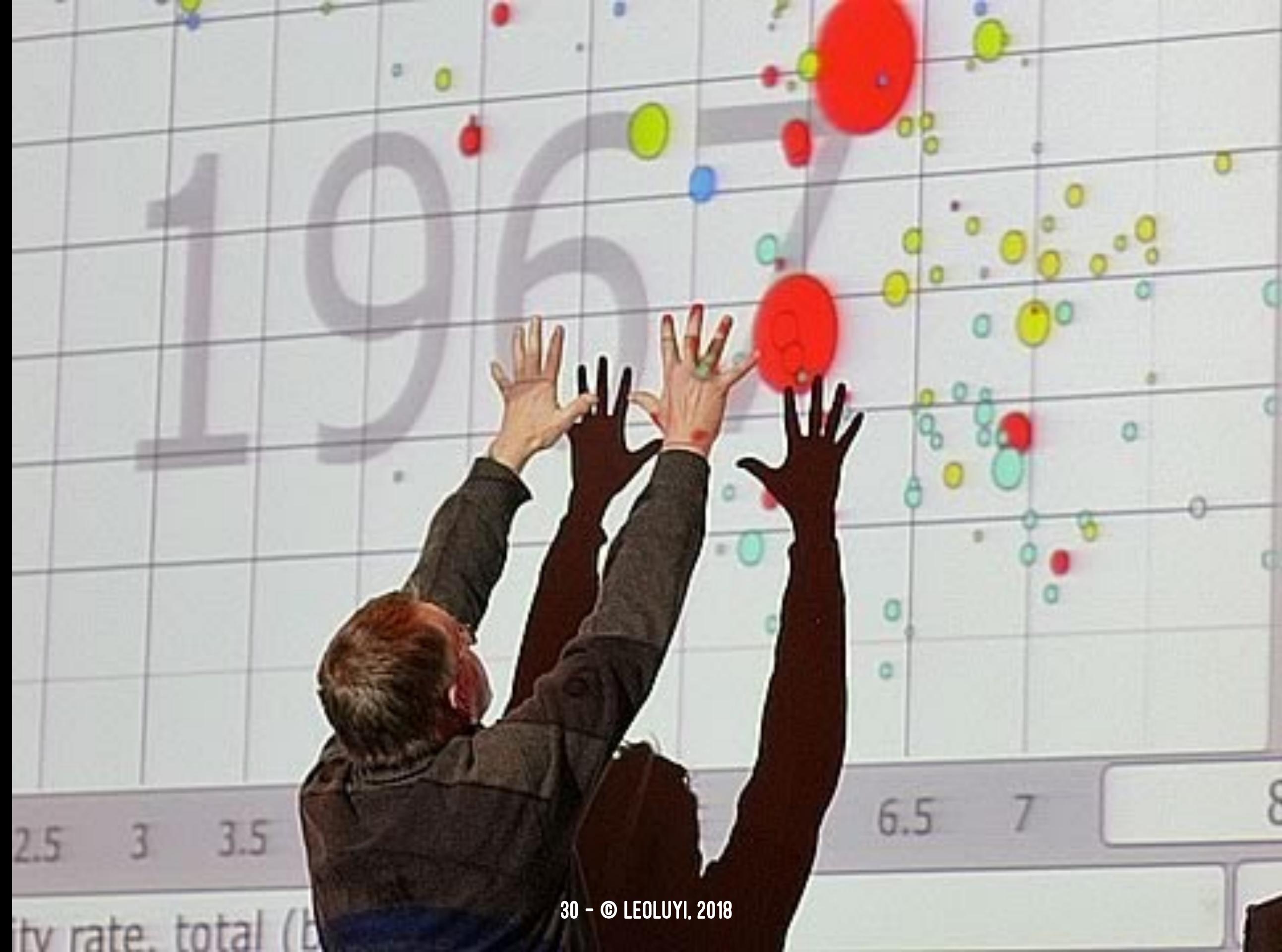
# VISUALISE THE MODEL

VIZ IS FUNDAMENTALLY A HUMAN  
ACTIVITY.

# MODEL + VIZ == GOOD BUSINESS

- > WE CANNOT SCALE GOOD VIZ
- > BUT MODELING CAN!

把你(複雜的)想法塞到別人腦中



# SUMMARY

# SUMMARY

› FROM DATA. TO MODEL. TO VIZ

# SUMMARY

- › FROM DATA, TO MODEL, TO VIZ
- › LEARN A VIZ LANGUAGE WITH ITS TOOLS

# SUMMARY

- › FROM DATA, TO MODEL, TO VIZ
- › LEARN A VIZ LANGUAGE WITH ITS TOOLS
  - › PLAY WITH YOUR DATA