

MÔN HỌC

**THỐNG KÊ ỨNG DỤNG (KC107)**

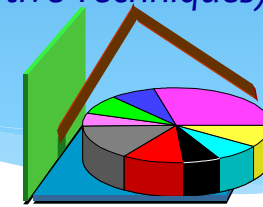
GIÁO VIÊN PHỤ TRÁCH

**ĐẶNG THỂ GIA**

Bộ môn Kỹ Thuật Xây Dựng  
Khoa Công Nghệ, Trường Đại Học Cần Thơ

Chương 2:

**KỸ THUẬT MÔ TẢ ĐỒ HỌA**  
(Graphical Descriptive Techniques)



Nội dung chương

1. **Giới thiệu**
2. **Các kiểu dữ liệu** (Types of data)
3. **Kỹ thuật đồ họa cho dữ liệu định lượng**  
(Graphical Techniques for Quantitative Data)
4. **Các dạng biểu đồ** (Pie Charts, Bar Charts,  
Line Charts)
5. **Sơ đồ phân tán** (Scatter Diagrams)

**GIỚI THIỆU**  
**INTRODUCTION**

## Giới thiệu

- **Thống kê mô tả** (Descriptive statistics) bao gồm việc sắp xếp, tóm tắt và trình bày dữ liệu nhằm diễn giải một cách có ý nghĩa và hỗ trợ quá trình ra quyết định.
- Các phương pháp thống kê mô tả sử dụng:
  - Các kỹ thuật đồ họa (Chương 2)
  - Các phép đo mô tả số (Chương 3)
- Những phương pháp này áp dụng cho:
  - Tổng thể/quần thể (the entire population)
  - Mẫu (the population sample)

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## CÁC KIỂU DỮ LIỆU

### TYPES OF DATA

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các kiểu dữ liệu (Types of data)

- **Biến (Variable)** – một đặc tính của tổng thể hoặc mẫu mà chúng ta quan tâm đến.
  - Lựa chọn sản phẩm trên kệ hàng
  - Chi phí đầu tư
  - Thời gian chờ xe buýt
- **Dữ liệu (Data)** – giá trị thực của biến
  - Dữ liệu định lượng (Quantitative data) – các quan sát dạng số (numerical observations)
  - Dữ liệu định tính (Qualitative data) – các quan sát thang đo

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Dữ liệu định tính & Dữ liệu định lượng (Categorical/Qualitative & Quantitative Data)

### Dữ liệu định lượng

<u>Tuổi – Thu nhập</u>	
55	75000
42	68000
.	.
.	.

<u>Tăng cân</u>
+10
+5
.
.

### Dữ liệu định tính

Với dữ liệu định tính, tất cả những gì chúng ta có thể làm là, tính toán tỷ lệ mà các dữ liệu rơi vào mỗi thang đo.

<u>Giảng viên</u>	<u>GS</u>	<u>PhóGS</u>	<u>Trợ giảng</u>	<u>Tổng</u>
15	25	5	15	60
25%	41.67%	8.33%	25%	

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Chọn kiểu dữ liệu

- Hiểu rõ về kiểu dữ liệu sẽ giúp chọn lựa được các kỹ thuật phù hợp để sử dụng.
- Đôi khi, đặc biệt là khi thực hiện các kỹ thuật phi tham số, cần phải biết liệu dữ liệu có thứ tự (rank) hay không.
- Kiểu phân tích cho phép đối với mỗi kiểu dữ liệu:
  - Dữ liệu định lượng – Các tính toán số học
  - Dữ liệu định tính – đếm số lượt quan sát của mỗi cấp thang đo
  - Dữ liệu có thứ tự – các tính toán dựa trên một trình tự (ordering process)

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Dữ liệu thời điểm & Dữ liệu chuỗi thời gian (Cross-Sectional & Time-Series Data)

- Dữ liệu cắt ngang/thời điểm (Cross-sectional Data) được thu thập tại một thời điểm nhất định: Khảo sát thị trường (quan sát các sở thích theo giới tính, tuổi,...), Điểm kiểm tra trong một khóa học của môn TKUD, Lương khởi điểm của sinh tốt nghiệp ngành XD,...
- Dữ liệu chuỗi thời gian (Time-series data) được thu thập qua các thời điểm liên tiếp: Giá vàng lúc đóng cửa hàng tuần, Lượng dầu thô nhập khẩu hàng tháng,...
- Dữ liệu cross-section chỉ ghi nhận sự kiện tại một thời điểm, trong khi dữ liệu time-series phải theo dõi đối tượng trong một thời gian có thể là nhiều năm tháng. Do đó, các nghiên cứu theo thời gian thường tốn kém hơn và công phu hơn các nghiên cứu tại một thời điểm.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

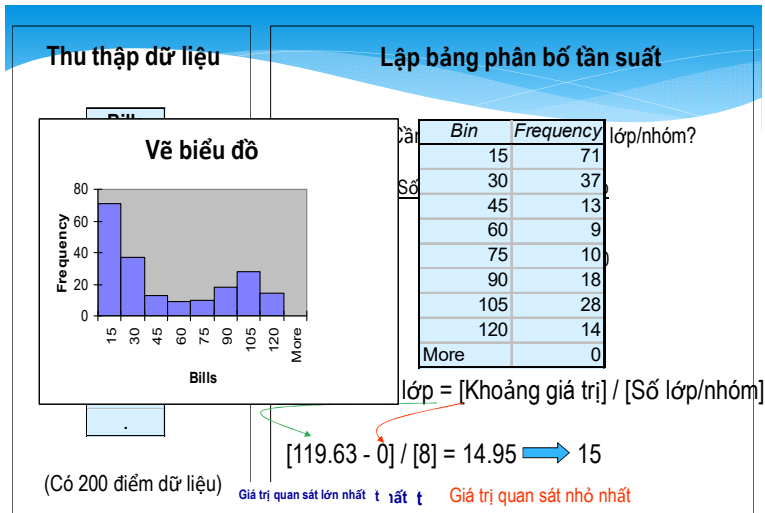
## KỸ THUẬT ĐỒ HỌA CHO DỮ LIỆU ĐỊNH LƯỢNG

### GRAPHICAL TECHNIQUES FOR QUANTITATIVE DATA

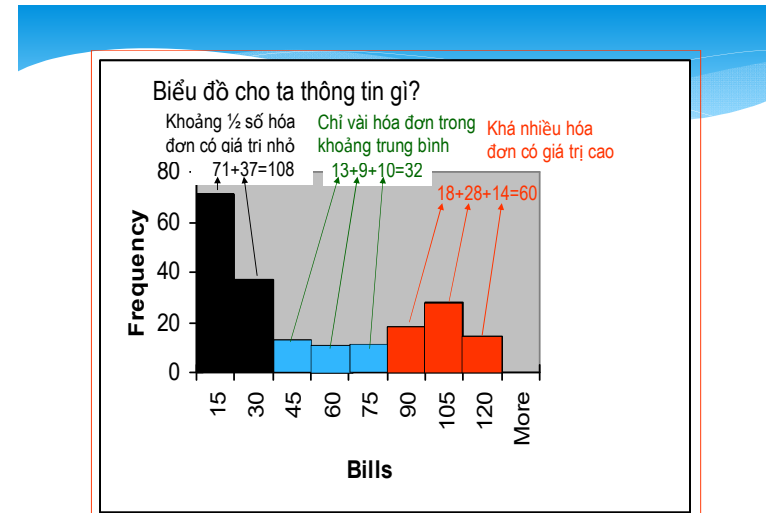
Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

- **Ví dụ:** Cung cấp thông tin liên quan đến hóa đơn hàng tháng của thuê bao mới trong tháng đầu tiên sau khi ký kết hợp đồng với một nhà mạng điện thoại.
- Các bước thực hiện:
  - Thu thập dữ liệu
  - Lập bảng phân phối tần suất
  - Vẽ biểu đồ

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Tần suất tương đối (Relative frequency)

- Đôi khi tần suất tương đối (tỉ lệ) sẽ thích hợp hơn khi tần suất để biểu diễn của số quan sát rơi vào trong mỗi nhóm.

$$\text{Tần suất tương đối của nhóm} = \frac{\text{Tần suất nhóm}}{\text{Tổng số quan sát}}$$

- Tần suất tương đối thường dùng khi:
  - Nghiên cứu tần suất tương đối của tổng thể
  - So sánh nhiều biểu đồ
  - Số lượng quan sát khác nhau trong các mẫu nghiên cứu

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

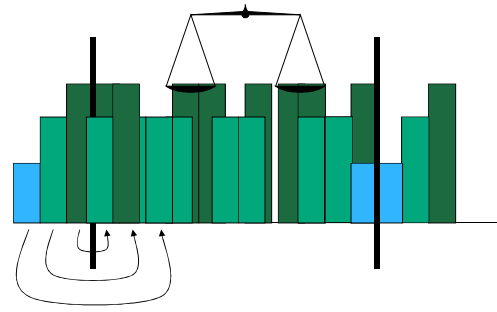
## Chiều rộng nhóm (Class width)

- Thông thường chiều rộng nhóm/lớp bằng nhau, nhưng đôi khi cũng cần sử dụng các chiều rộng nhóm/lớp không đồng đều.
- Độ rộng nhóm/lớp không đồng đều được sử dụng khi tần số liên quan đến những nhóm/lớp học quá thấp. Khi đó:
  - Vài nhóm được kết hợp lại với nhau để tạo thành một nhóm rộng hơn và "đông đúc" hơn.
  - Có thể tạo thành các nhóm mở ở 2 đầu (cao nhất và thấp nhất) của biểu đồ.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

# Hình dạng biểu đồ (Shapes of histogram)

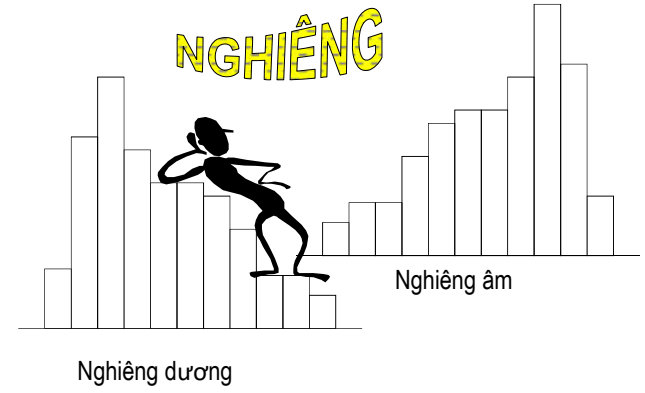
## ĐỐI XỨNG



Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

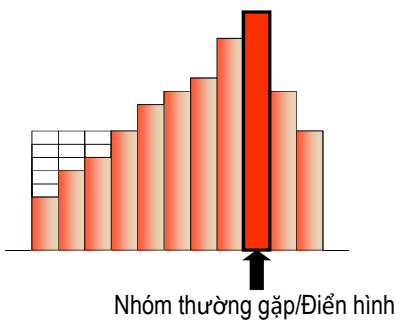
# Hình dạng biểu đồ (Shapes of histogram)

## NGHIÊNG



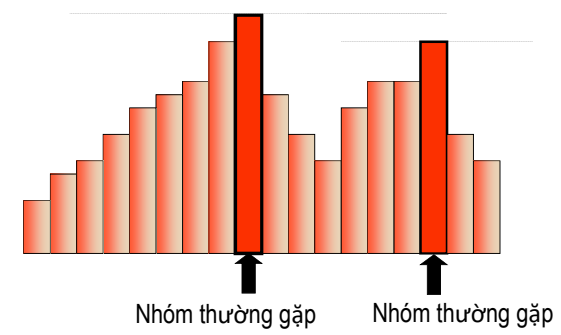
Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

# Biểu đồ nhóm thường gặp đơn (Unimodal Histogram)



Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

# Biểu đồ nhóm thường gặp đôi (Bimodal Histogram)

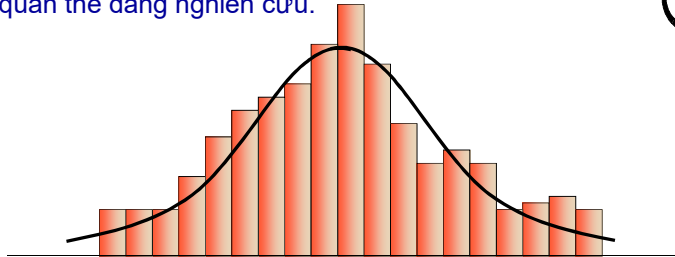


Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Biểu đồ hình chuông

(Bell shaped histogram)

- Nhiều kỹ thuật thống kê đòi hỏi tổng thể phải có hình chuông.
- Vẽ biểu đồ giúp xác minh hình dạng của quần thể đang nghiên cứu.

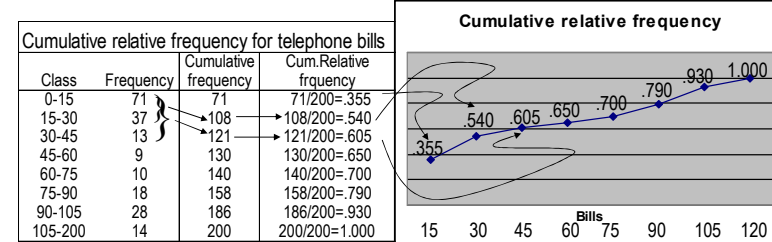


Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Biểu đồ hình cung/lũy tích

(Ogives)

- Hình cung (Ogives) biểu diễn một phân phối tần suất tương đối lũy tích
- Tiếp tục ví dụ trước



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Biểu diễn dạng Thân-Lá

(Stem-and-Leaf Display)

- Là biểu diễn dạng khoảng, rất hữu ích trong phân tích sơ bộ.
- Các sơ đồ Thân-Lá biểu diễn được giá trị của các quan sát nguyên thủy, trong khi biểu đồ (histogram) thường bị "mất" chúng.
- Biểu diễn Thân-Lá là công cụ trình bày dữ liệu định lượng ở dạng đồ họa, tương tự như một biểu đồ, để giúp hình dung hình dạng của phân bố.
- Các giá trị thường phải được làm tròn:  
-23.678758, -12.45, -3.4, 4.43, 5.5, 5.678, 16.87, 24.7, 56.8

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Biểu diễn dạng Thân-Lá



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Tạo sơ đồ Thân-Lá (Creating Stem-and-Leaf Display)

Quan sát dữ liệu trong bảng sau

19.1	19.8	18.0	19.2	19.5	17.3	20.0	20.3
19.6	18.5	18.1	19.7	18.4	17.6	21.2	20.6
22.2	19.1	21.1	19.3	20.8	21.2	21.0	18.7
19.9	18.7	22.1	17.2	18.4	21.4		

Xác định cái gì là Lá & cái gì là Thân (Có nhiều cách).

Ví dụ:

- Chữ số bên trái dấu thập phân là Thân.
- Chữ số bên phải dấu thập phân là Lá.

19 . 1  
Thân Lá

## Tạo sơ đồ Thân-Lá (Creating Stem-and-Leaf Display)

Liệt kê "thân" vào cột theo thứ tự tăng dần  
Đặt "lá" và cùng hàng với "thân" của nó.

19.1 19.8 18.0 19.2 19.5 17.3  
19 19 18 19 19 17

### Kết luận từ Sơ đồ Thân-Lá:

- Giá trị thường gặp là 19.
- Các giá trị quan sát nằm trong khoảng 17.2 - 22.2
- Hầu hết các giá trị nằm trong khoảng giữa 18.0 và 20.0
- Hình dạng của phân bố không đối xứng.
- 1/2 số giá trị quan sát nằm dưới 19.5 và 1/2 nằm trên giá trị này.

Biểu diễn hoàn chỉnh

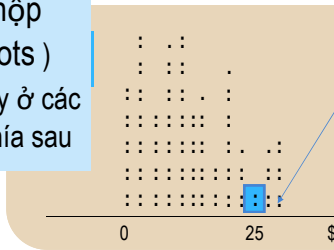
Stem	Leaf
17	623
18	4705147
19	1983627571
20	038
21	12204
22	12

## Đồ thị chấm (Dot Plot/Chart)

Đây là dạng đồ thị tương tự biểu đồ histogram, trong đó:

- Trục ngang được chia thành nhiều nhóm/lớp
- Các giá trị quan sát được biểu diễn bằng các dấu chấm

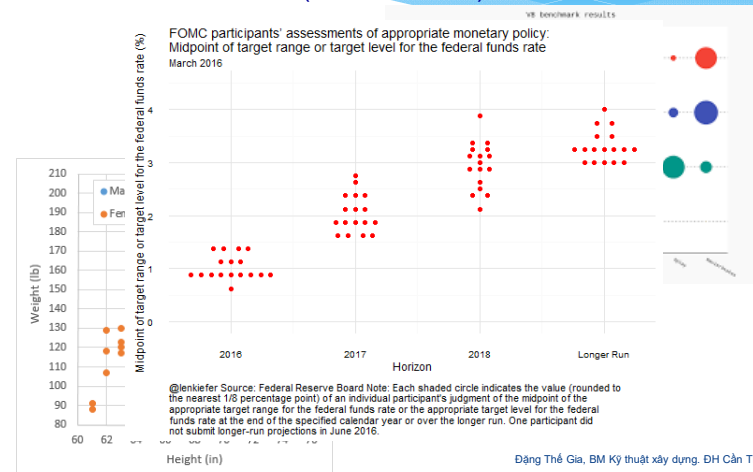
• Đồ thị hộp  
(Box Plots)  
Trình bày ở các  
slides phía sau



Có 10 khoảng trống cho mỗi \$25. Mỗi khoảng trống "trị giá" \$2.5

Có 2 dấu chấm ở vị trí \$25 đại diện cho 2 quan sát rơi vào giữa khoảng [25-1.25, 25+1.25]

## Đồ thị chấm (Dot Plot/Chart)

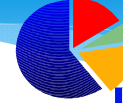


## CÁC DẠNG BIỂU ĐỒ

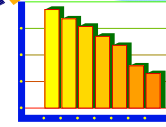
### PIE CHART, BAR CHART & LINE CHART

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

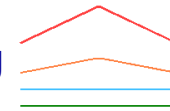
### Biểu đồ bánh



### Biểu đồ thanh



### Biểu đồ đường thẳng



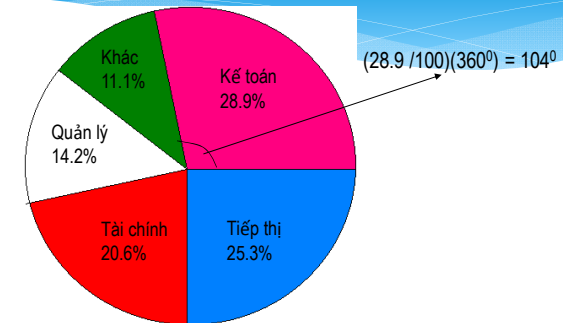
- Sử dụng chủ yếu cho dữ liệu định tính.
- Những công cụ đồ họa này là thích hợp nhất khi dữ liệu thô có thể được phân loại tự nhiên một cách có ý nghĩa.

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Biểu đồ bánh/tròn (Pie/Circle Chart)

- Biểu đồ hình tròn là một công cụ phổ biến dùng để biểu thị tỷ lệ của sự xuất hiện cho dữ liệu danh nghĩa.
- **Ví dụ 2.3**
  - Phòng Công Tác Sinh Viên của CTU muốn xác định tổng thể lĩnh vực việc làm của năm rồi.
  - Dữ liệu đếm đã được thu thập, và số lần xuất hiện đã được ghi lại cho từng lĩnh vực.
  - Những dữ liệu đếm này được chuyển thành tỷ lệ và kết quả được trình bày trong một biểu đồ tròn.

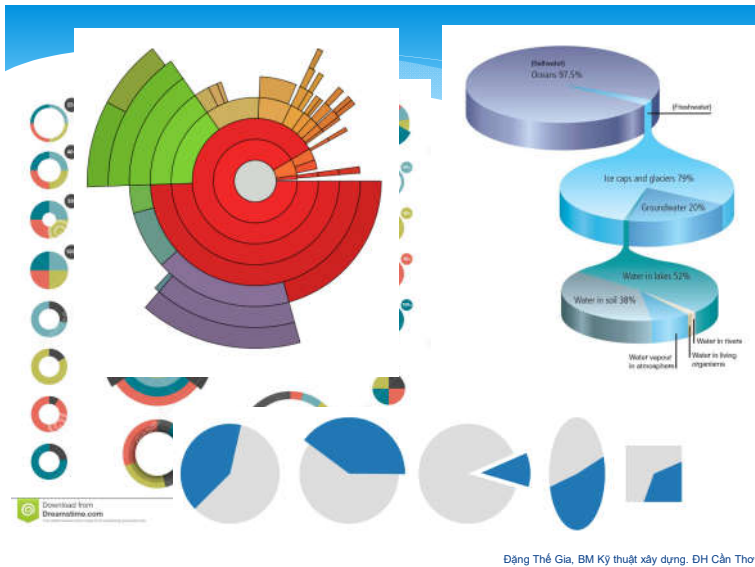
Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



- Biểu đồ hình tròn là một vòng tròn, được chia nhỏ thành một số miếng tương ứng với các lĩnh vực việc làm.
- Kích thước của mỗi miếng tỷ lệ với tỷ lệ phần trăm tương ứng với số đếm được của mỗi lĩnh vực mà nó đại diện.

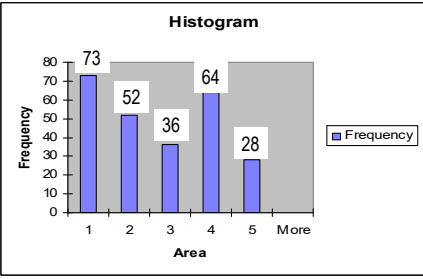
Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ





## Biểu đồ Thanh (Bar Chart)

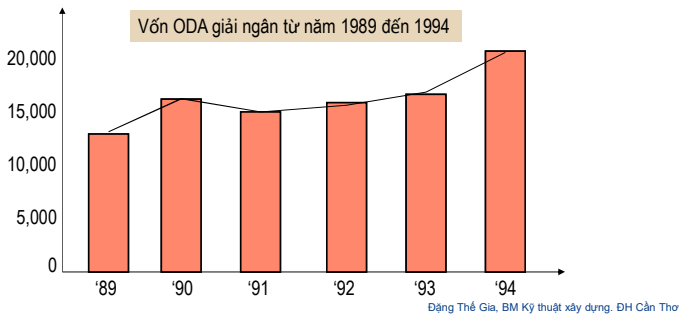
- Biểu đồ thanh cung cấp một giải pháp khác cho biểu đồ bánh
- Tần suất (tần suất tương đối) của mỗi nhóm được biểu diễn bằng một thanh đứng.
- Ví dụ trên được biểu diễn dạng biểu đồ thanh



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

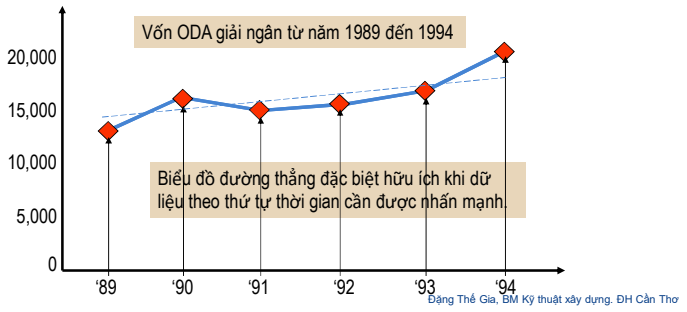
## Biểu đồ Thanh (Bar Chart)

Dùng biểu đồ thanh khi trật tự của dữ liệu định lượng được trình bày có ý nghĩa.

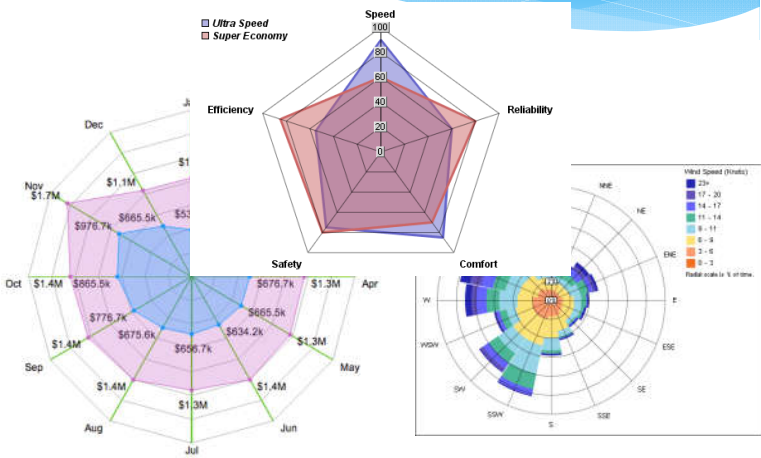


## Biểu đồ đường thẳng (Line Chart)

- Chấm các giá trị ở phía trên mỗi vạch của trục ngang
- Nối các dấu chấm bằng đoạn thẳng lần lượt theo thứ tự

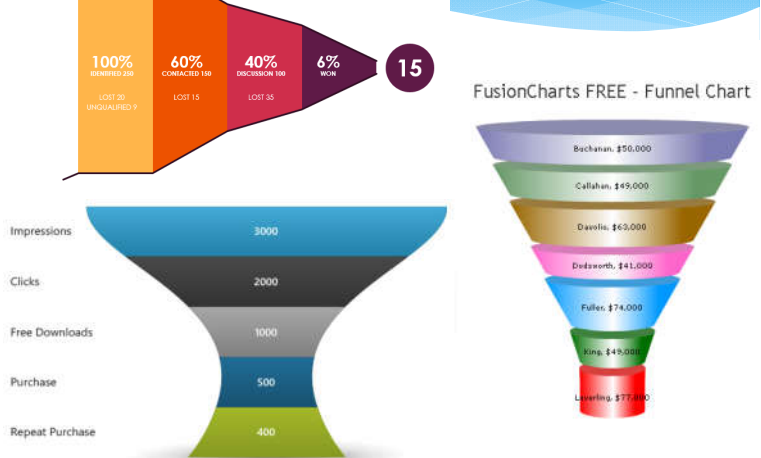


### Biểu đồ radar (Radar Chart)



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

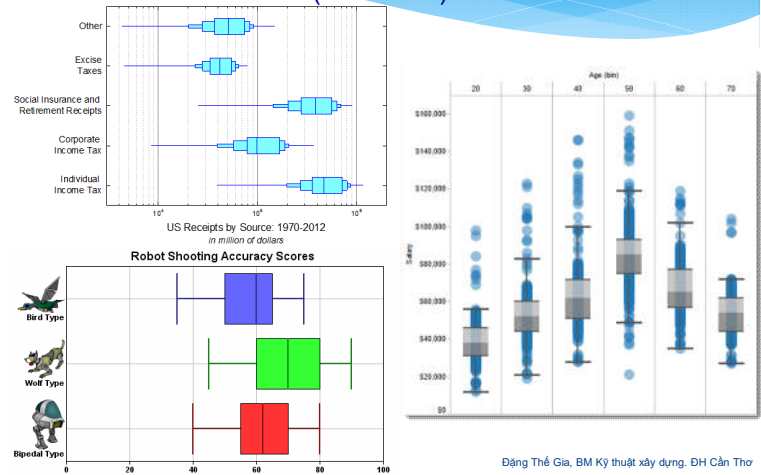
### Biểu đồ ống khói (Funnel Chart)



FusionCharts FREE - Funnel Chart

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

### Biểu đồ hộp (Box Chart)



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

### SƠ ĐỒ PHÂN TÁN SCATTER DIAGRAM

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Sơ đồ phân tán (Scatter Diagram)

- Thông thường chúng ta quan tâm đến mối quan hệ giữa hai biến định lượng.

### Ví dụ 2.4

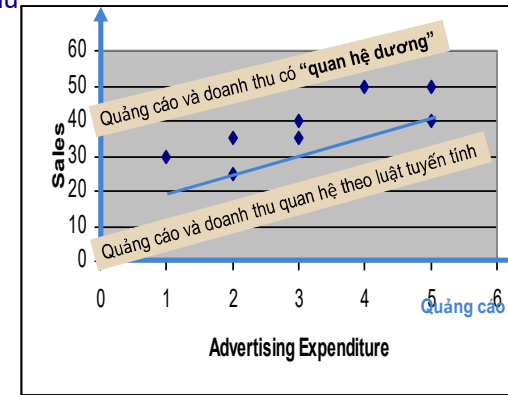
- Một doanh nghiệp quảng cáo trên các mức chi tiêu khác nhau để đánh giá hiệu quả quảng cáo.
- Dữ liệu quan sát được trong bảng dưới đây. Mỗi cặp bao gồm chi tiêu quảng cáo và doanh thu hàng tháng.

Advert	Sales
1	30
3	40
5	40
4	50
2	35
5	50
3	35
2	25

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

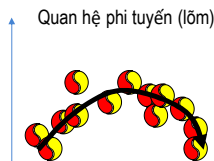
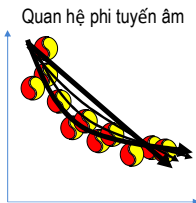
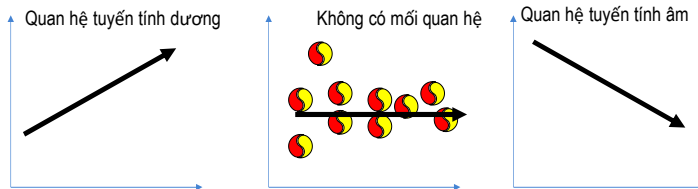
Sơ đồ phân tán biểu diễn mối quan hệ giữa chi phí quảng cáo và doanh thu

Advert	Sales
1	30
3	40
5	40
4	50
2	35
5	50
3	35
2	25



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

## Các quan hệ điển hình



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ