

Lựa chọn cách trực quan hóa dữ liệu hiệu quả

Nguyễn Thị Bích Thủy

Email: nbthuy2001@gmail.com

Tel: 0981 365 780

Nội dung

1. Chú thích
2. Bảng
3. Biểu đồ
4. Một số loại biểu đồ khác
5. Những điều nên tránh
6. Kết luận

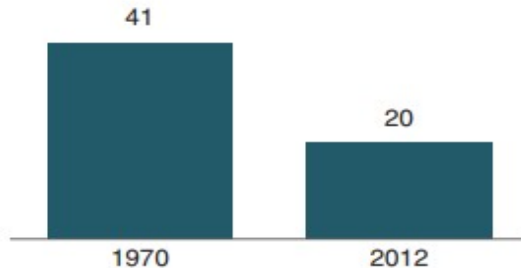
I. Chú thích

- Khi chỉ có một hoặc hai số trong dữ liệu
- Không nên:

Nên:

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

FIGURE 2.3 Stay-at-home moms simple text makeover

II. Tables

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

FIGURE 2.4 Table borders

- Dễ dàng nắm bắt thông tin.
- Cân nhắc dùng bảng khi thuyết trình, người xem dễ tập trung vào bảng mà quên mất ý tưởng của người nói.
- Viền bảng nên để nhạt hoặc xóa bỏ, dữ liệu mới là thứ cần tập trung và nên để vào vị trí trung tâm.

II. Tables

- Heatmap là một dạng của tables
- Tập trung biểu diễn thông tin qua các tín hiệu thị giác.

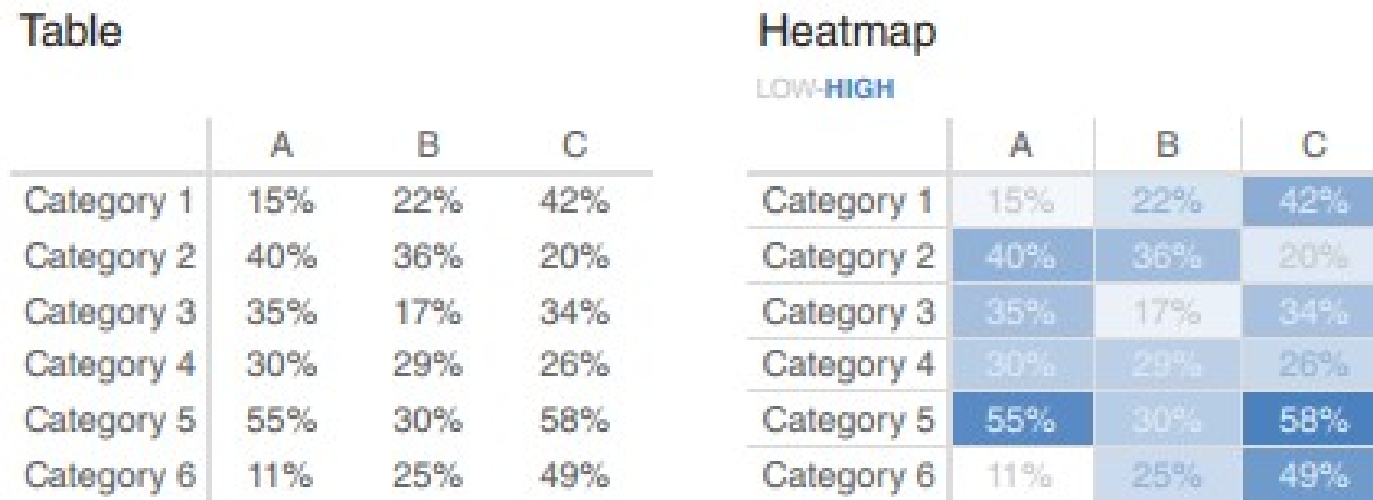


FIGURE 2.5 Two views of the same data

III. Graphs

Graphs gồm :

- Points
- Lines
- Bars
- Areas

III. Graphs

1. Points

- Biểu thị mối quan hệ giữa hai biến

Cost per mile by miles driven

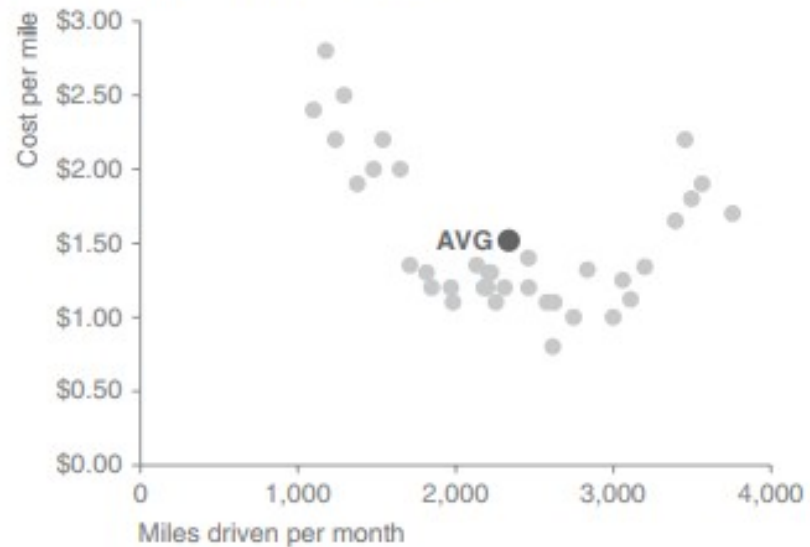


FIGURE 2.6 Scatterplot

Cost per mile by miles driven

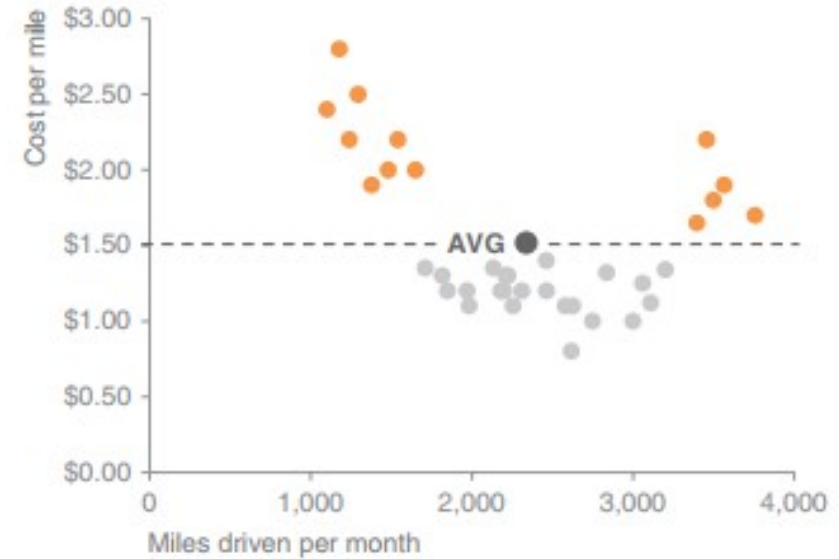


FIGURE 2.7 Modified scatterplot

III. Graphs

2. Lines

- Biểu diễn dữ liệu dạng liên tục
- Gồm 2 dạng: Line graph và Slope graph

2.1 Line graph

- Thang đo trên các trục nhai nhất quán theo các khoảng thời gian

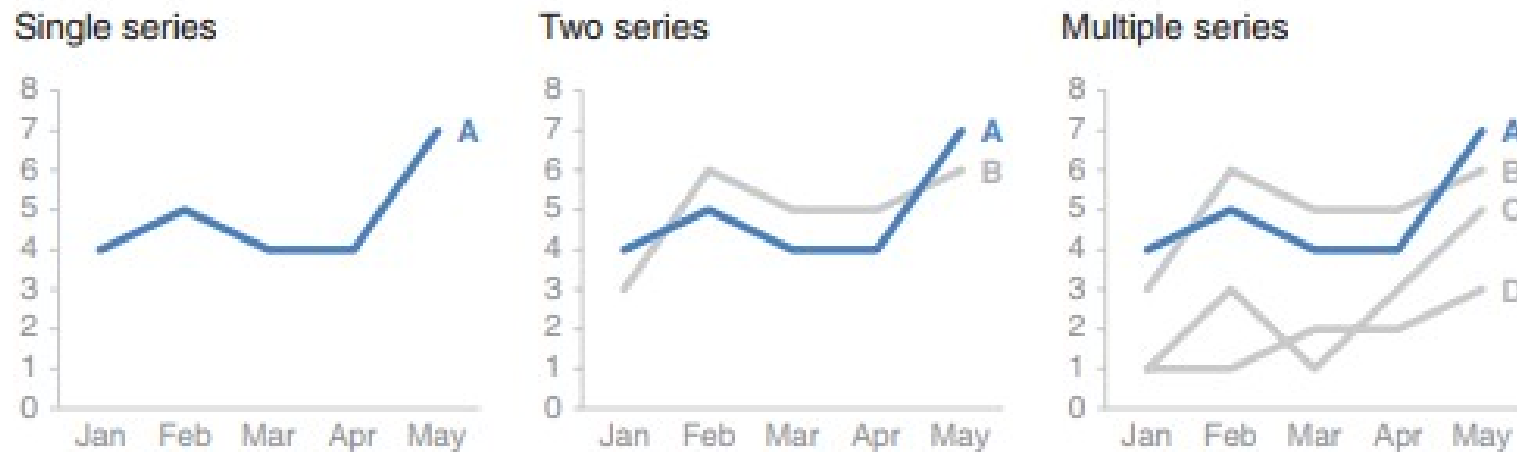


FIGURE 2.8 Line graphs

III. Graphs

2. Lines

- Biểu diễn dữ liệu dạng liên tục

2.2 Slope graph

- Nhanh chóng chỉ ra mối quan hệ tăng giảm hay sự khác biệt giữa các biến

Employee feedback over time

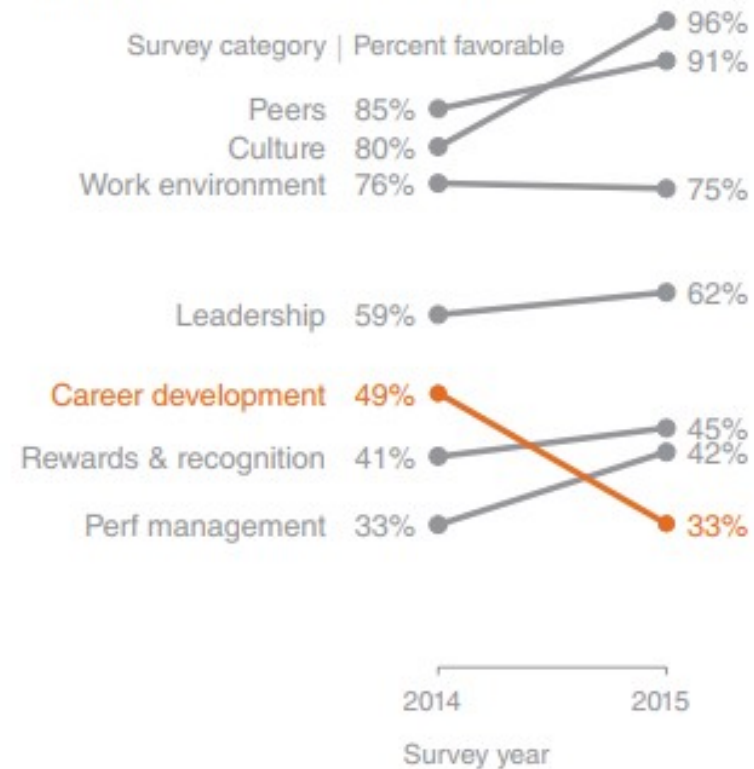


FIGURE 2.11 Modified slopegraph

III. Graphs

3. Bars

- Giúp người xem dễ dàng cảm nhận dữ liệu thông qua trực quan: giá trị lớn nhất, nhỏ nhất, mức độ gia tăng khác nhau ...

- Lưu ý:

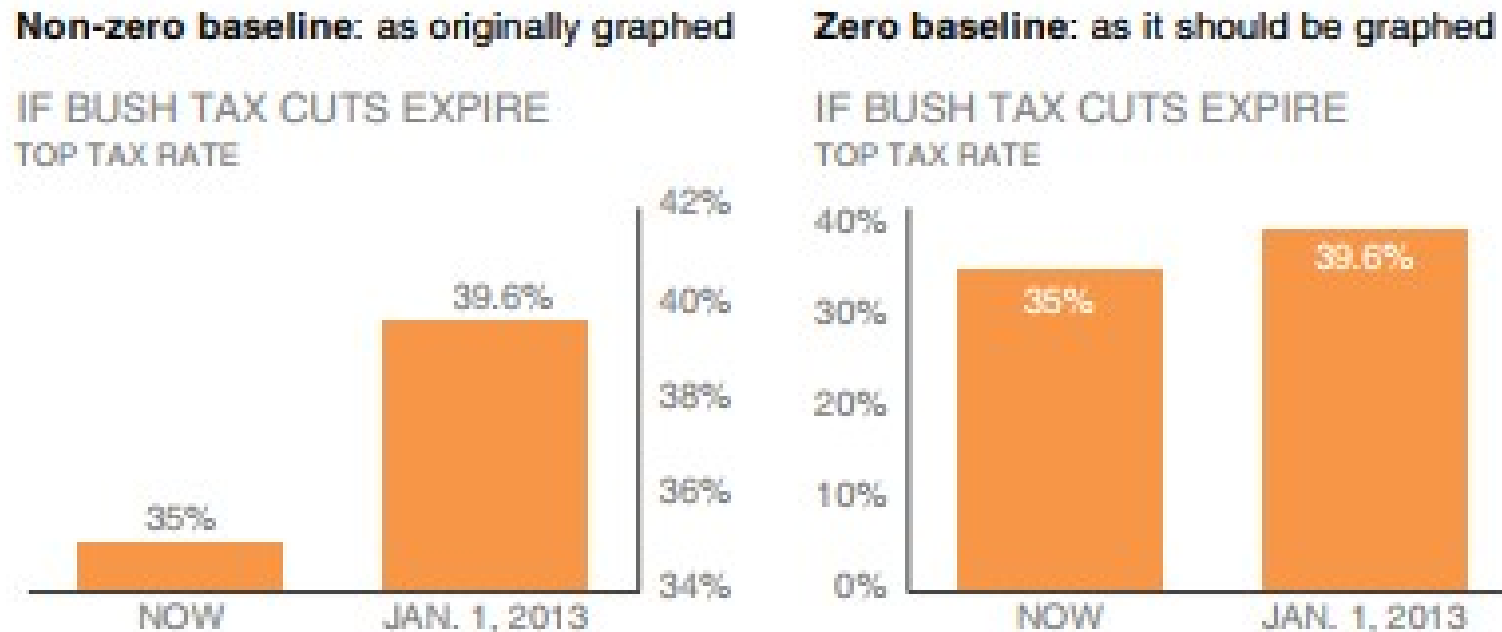


FIGURE 2.13 Bar charts must have a zero baseline

III. Graphs

3. Bars

- Gồm 5 dạng: *Vertical bar chart*, *Stacked vertical bar chart*, *Waterfall chart*
Horizontal bar chart, *Stacked horizontal bar chart*

- Lưu ý:



FIGURE 2.14 Bar width

III. Graphs

3. Bars

3.1 Vertical bar chart

- Biểu đồ thanh dọc
- Thêm nhiều dữ liệu, nhiều thanh dọc, khán giả khó tập trung
- Chú ý vị trí tương đối của các nhóm

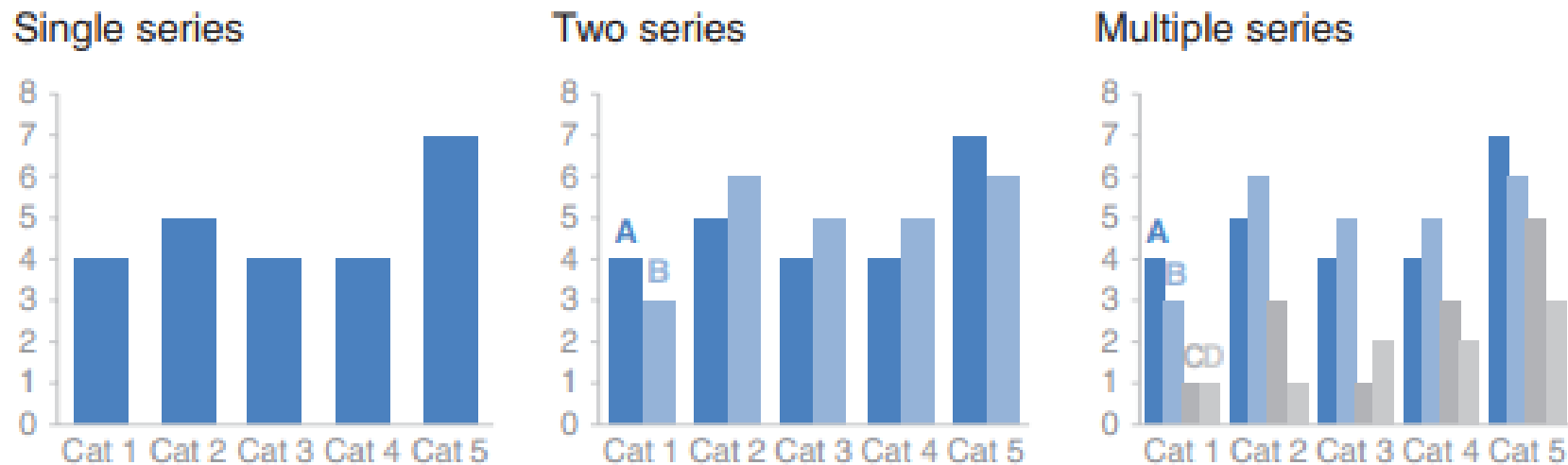


FIGURE 2.15 Bar charts

III. Graphs

3. Bars

3.2 Stacked vertical bar chart

- Các trường hợp có thể sử dụng Stacked vertical bar chart khá hạn chế
- So sánh tổng số của các danh mục khác nhau và từng thành phần con bên trong
- Thành phần con nằm kế bên trục Ox giữa các danh mục để so sánh nhất

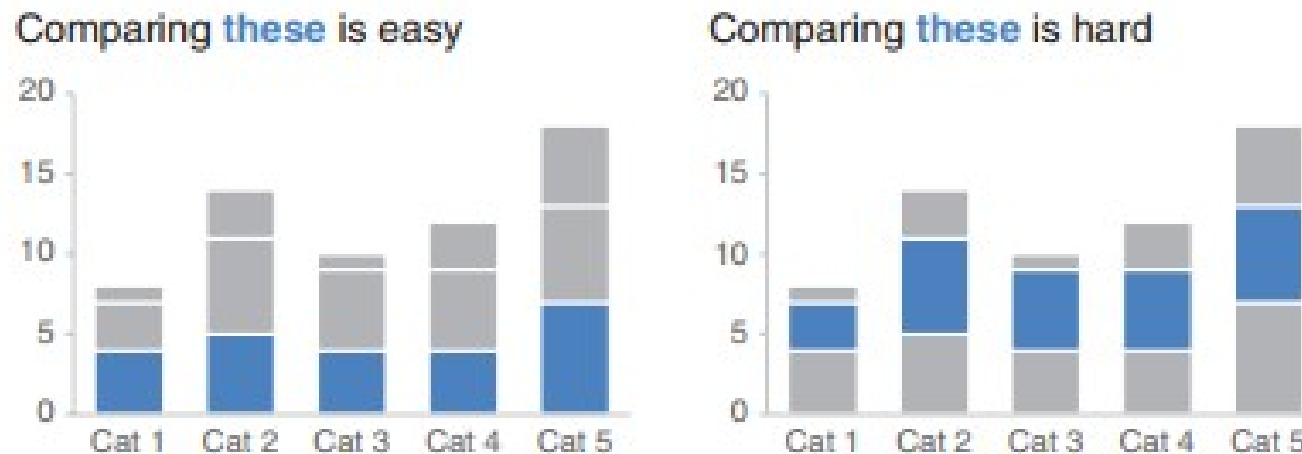


FIGURE 2.16 Comparing series with stacked bar charts

III. Graphs

3. Bars

3.3 Waterfall chart

- Sử dụng để tập trung vào từng giai đoạn, hoặc sử dụng để biểu diễn vị trí ban đầu, sự tăng giảm, và kết quả ở giá trị cuối cùng

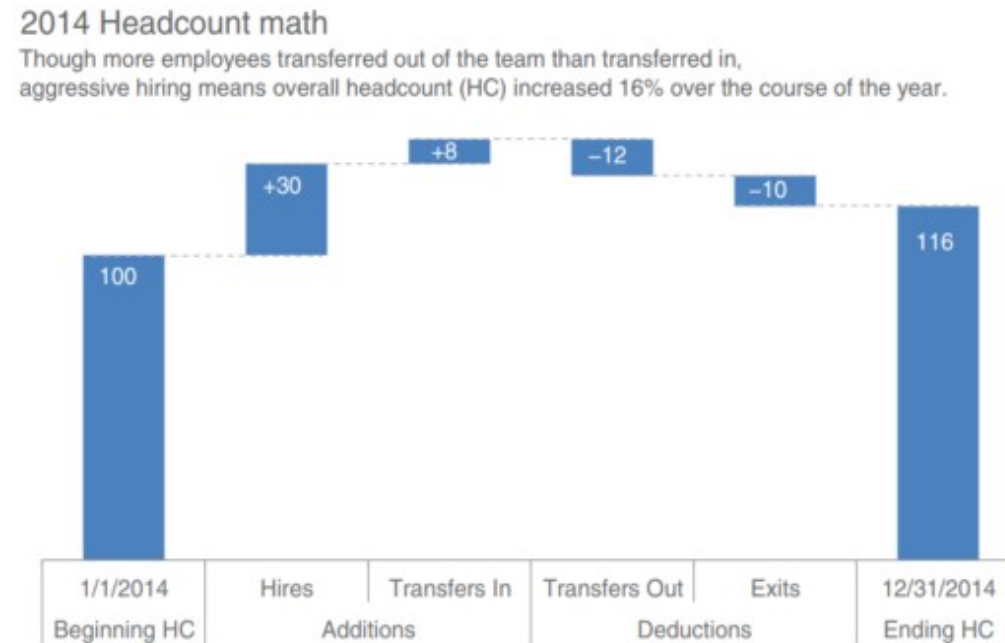


FIGURE 2.17 Waterfall chart

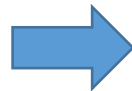
III. Graphs

3. Bars

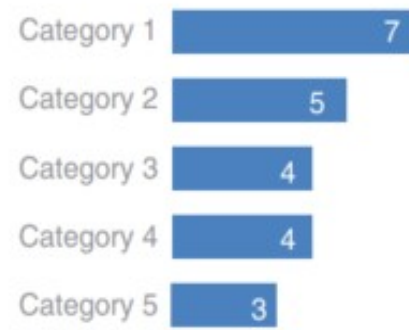
3.4 Horizontal bar chart (Biểu đồ thanh ngang)

- Tốt nhất để biểu diễn dữ liệu định tính
- Phù hợp khi tên biến định tính/ tên danh mục dài
- Dễ quan sát từ trái sang phải

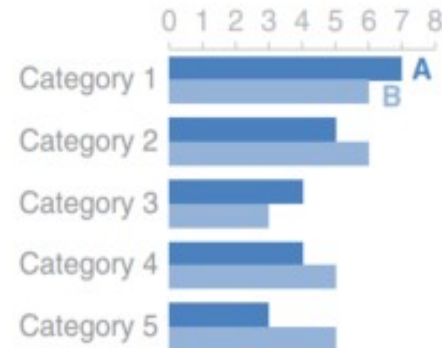
NOTE: Sắp xếp tên danh mục nếu cần



Single series



Two series



Multiple series

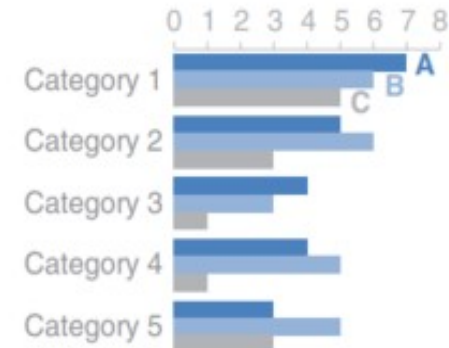


FIGURE 2.18 Horizontal bar charts

III. Graphs

3. Bars

3.5 Stacked horizontal bar chart (Biểu đồ cột chồng ngang)

- So sánh từng phần trên tổng số
- Phù hợp khi biểu diễn dữ liệu khảo sát được thu thập theo thang đo Likert (thang đo thường được sử dụng trong khảo sát trong khoảng từ Rất không đồng ý đến Hoàn toàn đồng ý)

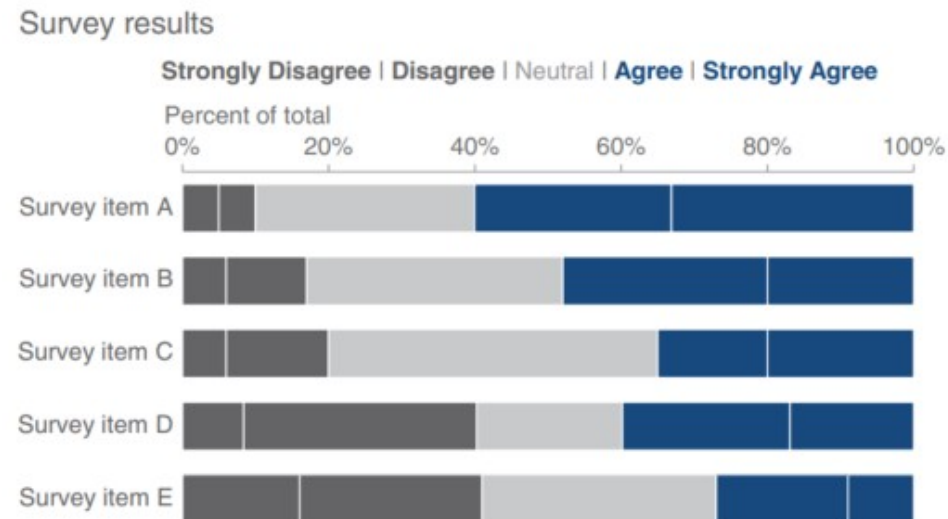


FIGURE 2.19 100% stacked horizontal bar chart

III. Graphs

4. Area

- Ít được sử dụng.
- Phù hợp khi biểu diễn số liệu có độ lớn khác biệt
- Đưa ra cái nhìn tổng quan nhanh về kích thước tương đối mà không sử dụng thang đo

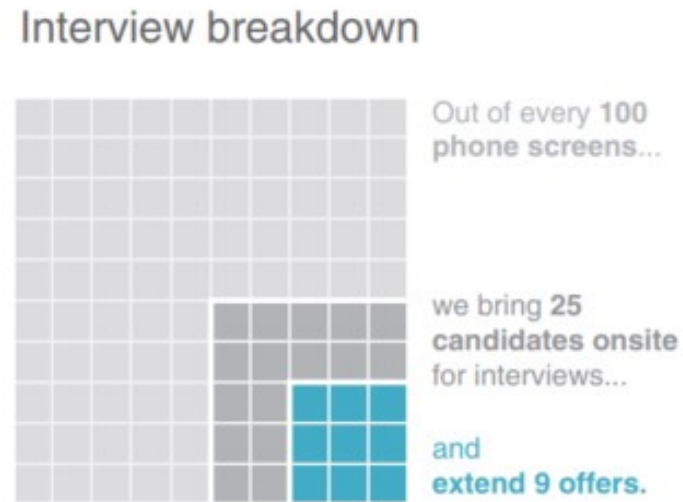


FIGURE 2.20 Square area graph

IV. Các loại biểu đồ khác

- Những loại biểu đồ trên đáp ứng phần lớn nhu cầu sử dụng
- Trả lời câu hỏi: Ai là khán giả của bạn? Bạn muốn họ biết hay làm gì?
→ Chọn loại biểu đồ phù hợp.



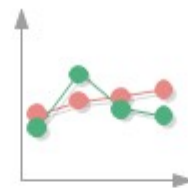
Pie



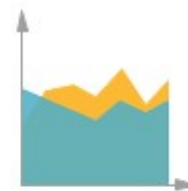
Bar



Column



Line



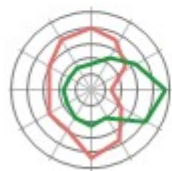
Area



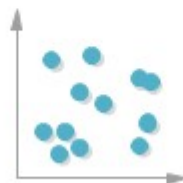
Doughnut



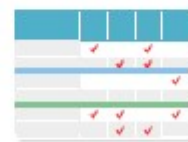
Bubble Chart



Spider and Radar



Scatter



Comparison Chart



Stacked bar chart



Gauges

V. Những điều nên tránh

1. Biểu đồ quạt không tốt

- Có thể khó quan sát, so sánh giữa các phần
- Thay bằng biểu đồ thanh ngang + sắp xếp → dễ dàng đánh giá, so sánh, tìm giá trị lớn nhất.

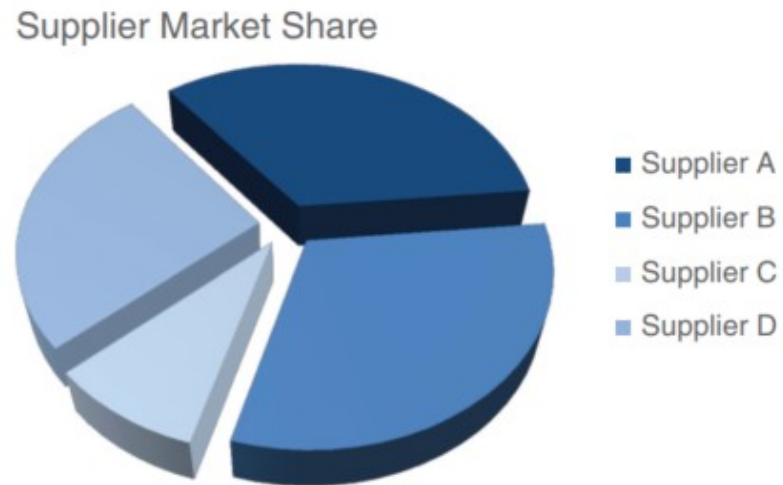


FIGURE 2.21 Pie chart

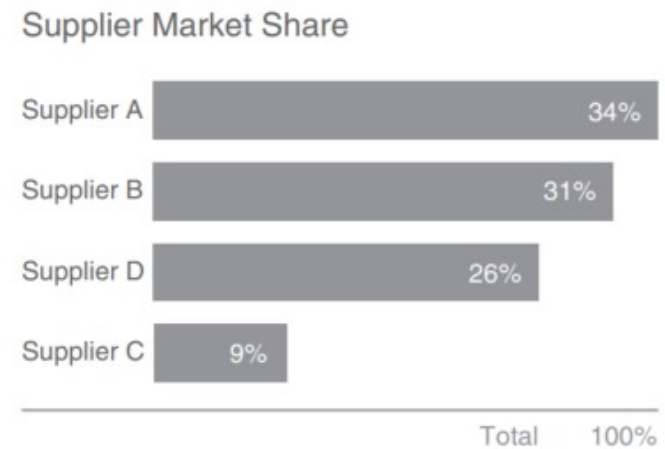
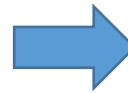


FIGURE 2.23 An alternative to the pie chart

V. Những điều nên tránh

1. Biểu đồ quạt không tốt

- Biểu đồ bánh Donut: tương tự biểu đồ quạt → khó đánh giá, so sánh.

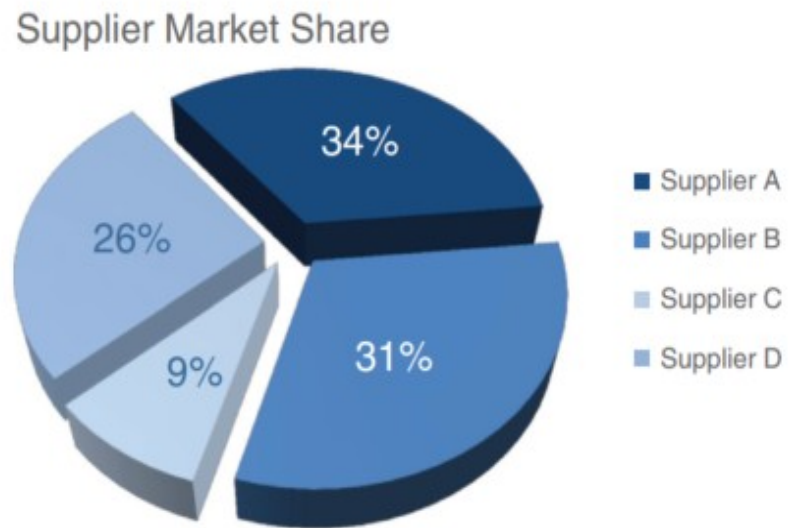


FIGURE 2.22 Pie chart with labeled segments

The donut chart



FIGURE 2.24 Donut chart

V. Những điều nên tránh

2. Đừng bao giờ sử dụng biểu đồ 3D

- Phức tạp, dễ gây nhầm lẫn trong số liệu → khó khăn trong so sánh dữ liệu.
- Chỉ sử dụng khi thực sự biểu diễn dữ liệu trong không gian 3 chiều.

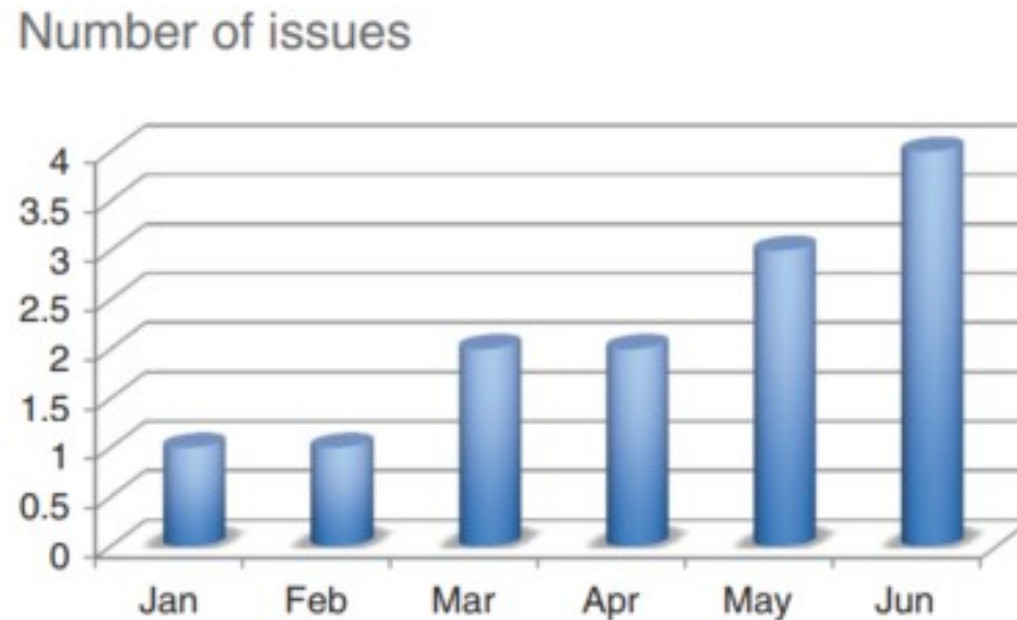


FIGURE 2.25 3D column chart

V. Những điều nên tránh

3. Hai trục y thường không phải là một ý tưởng hay

- Phải xác định biểu đồ nào ứng với trục nào → mất thời gian
- 2 giải pháp thay thế: **gán nhãn trực tiếp cho từng điểm dữ liệu** hoặc **tách riêng biểu đồ**



FIGURE 2.26 Secondary y-axis



Alternative 1: label directly



FIGURE 2.27 Strategies for avoiding a secondary y-axis

Alternative 2: pull apart vertically



V. Kết luận

- Biết cách sử dụng những loại biểu đồ thông dụng nhất.
- Không có một cách biểu diễn chính xác tuyệt đối nào, có nhiều cách biểu diễn cho cùng 1 mục đích.
- **Quan trọng nhất là xác định rõ mục tiêu:** Bạn muốn người xem biết thêm về điều gì? → Chọn kiểu biểu đồ phù hợp với mục tiêu → Loại biểu đồ nào giúp người xem dễ dàng quan sát và dễ hiểu nhất.
- Có thể nhờ bạn bè đồng nghiệp review: tập trung nhất vào phần nào? quan sát thấy gì? có câu hỏi gì? → Đánh giá, cải tiến.

