

# Effective Data Visualisation & Storytelling

Khoa Khoa học Dữ liệu và Trí tuệ nhân tạo (FDA)  
Trường Công nghệ, Trường Đại học Kinh tế Quốc dân

VNPT AI  
Tập đoàn Bưu chính Viễn thông Việt Nam VNPT





# Learning Objectives

1

Nắm được quy trình 6 bước trực quan hóa dữ liệu hiệu quả

2

Biết cách tư duy và đặt câu hỏi khi tiến hành biểu diễn dữ liệu

3

Hiểu được các nguyên tắc Gestalt và Tufte khi biểu diễn dữ liệu

4

Biết cách lựa chọn visual và loại bỏ nhiễu để việc biểu diễn dữ liệu chính xác và hiệu quả nhất

5

Thực hành xây dựng kịch bản/câu chuyện và mock presentation với một mẫu visual



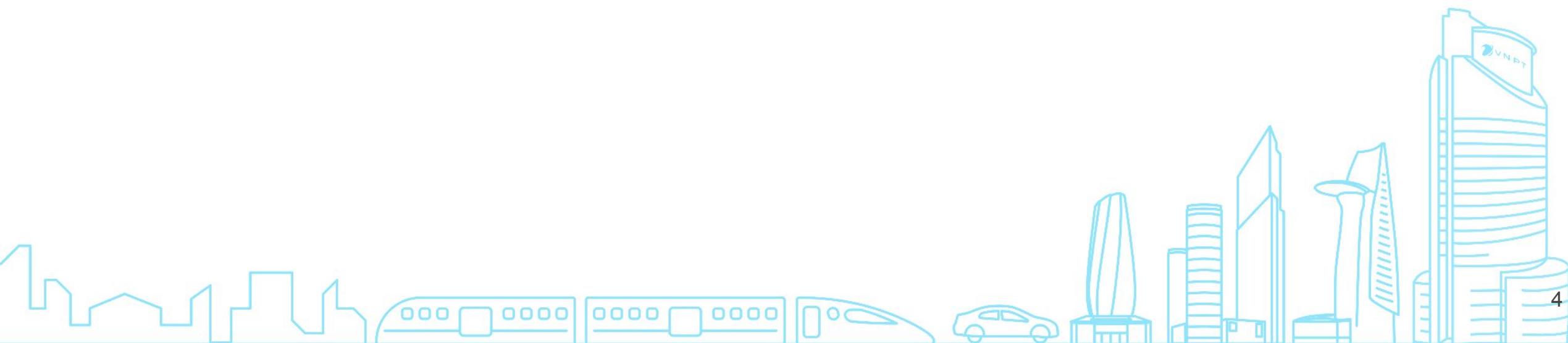
1

# Quy trình biểu diễn dữ liệu





# What is Data Visualisation and Why?



# Biểu diễn/ Trực quan hóa dữ liệu



## Khái niệm

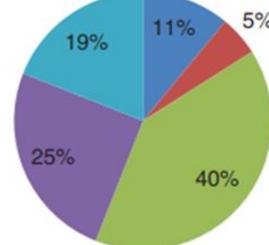
Data Visualisation: quá trình **chuyển đổi dữ liệu** thành các định dạng **đồ họa** như biểu đồ, bản đồ, quy trình, etc. Phương pháp này giúp người xem **nắm bắt thông tin phức tạp một cách chính xác, nhanh chóng, hiệu quả**.

Làm cho dữ liệu **dễ hiểu** và **gợi mở những thông tin** chi tiết có thể không dễ dàng nhận thấy khi xem dữ liệu thô.

## Survey Results

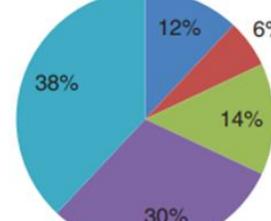
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



## Pilot program was a success

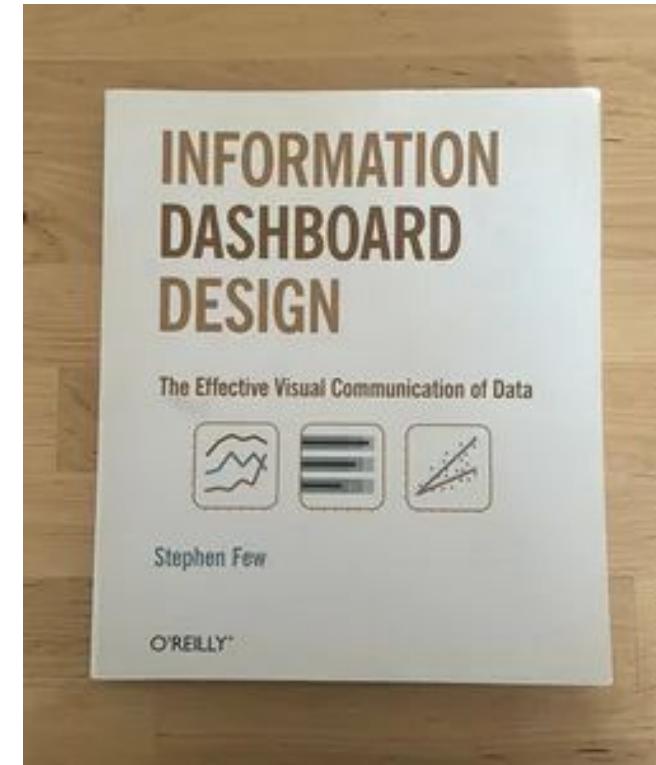
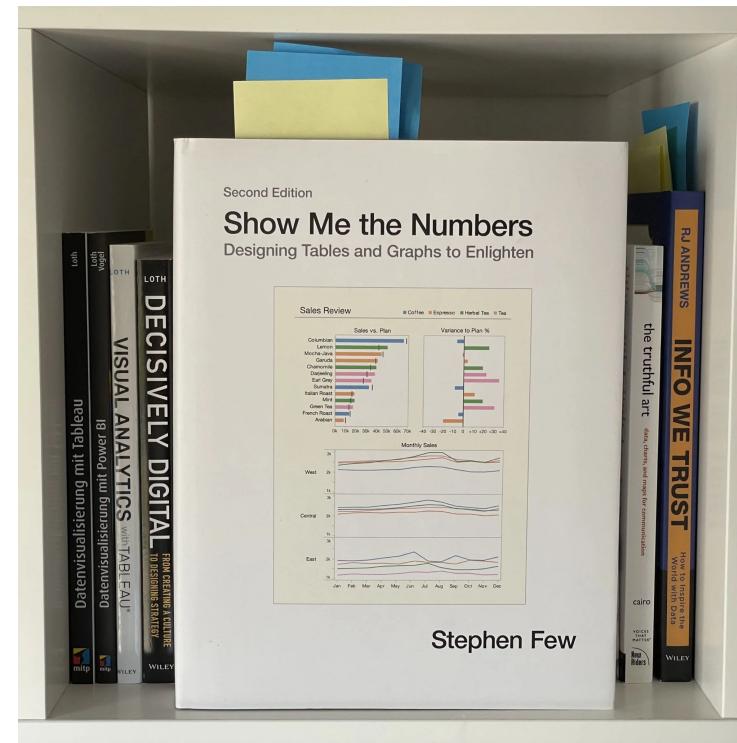
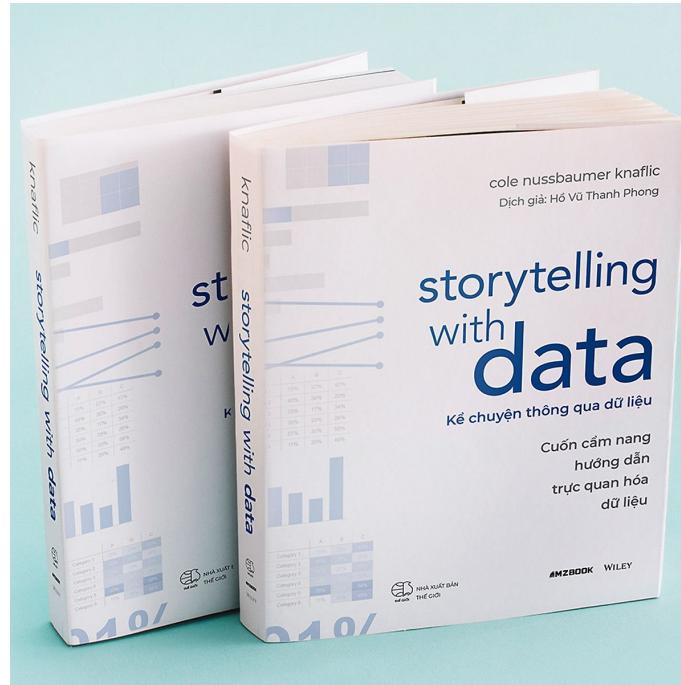
How do you feel about science?

**BEFORE** program, the majority of children felt just **OK** about science.



Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).





## Quy trình trực quan hóa dữ liệu



- 1 Hiểu bối cảnh, mục tiêu thông tin cần truyền tải
- 2 Lựa chọn biểu đồ phù hợp để biểu diễn thông tin
- 3 Loại bỏ các yếu tố gây nhiễu
- 4 Làm nổi bật các yếu tố quan trọng cần truyền tải
- 5 Hoàn thiện, Kiểm tra và Điều chỉnh
- 6 Storytelling (Kết hợp lời dẫn để tạo nên một câu chuyện thuyết phục)

## Quy trình trực quan hóa dữ liệu

1

### Hiểu bối cảnh thông tin cần truyền tải

- Người nghe/ người ra quyết định là ai?
- Họ muốn biết thông tin gì/ cần gì để ra quyết định
  
- Nhà quản trị cấp cao/ Strategic
  - Thông tin về xu hướng, nhu cầu để xây dựng kế hoạch dài hạn, ra quyết định chiến lược
- Nhà quản trị cấp trung/ Tactical
  - Thông tin trung-ngắn hạn về doanh thu từng dòng sản phẩm, phản hồi khách hàng, hiệu quả của các chiến dịch
- Nhà quản trị cấp cơ sở/ Operational
  - Dữ liệu hàng tồn kho, doanh số bán của từng nhân viên để tối ưu quy trình, quản lý năng suất và xử lý vấn đề phát sinh.



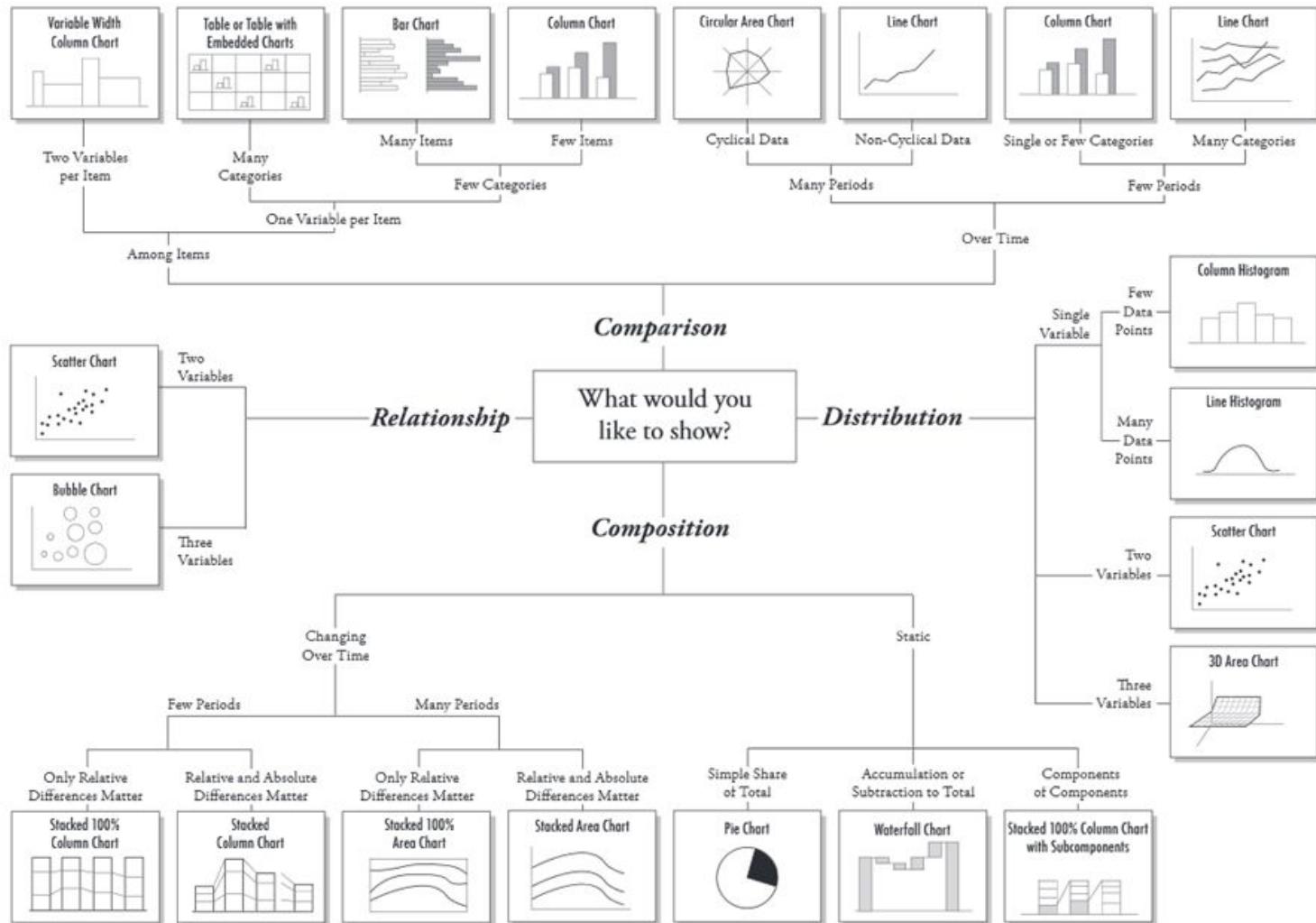
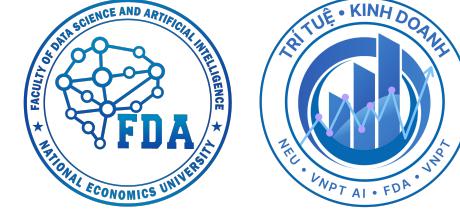


2

## Lựa chọn biểu đồ phù hợp



# Câu hỏi là gì

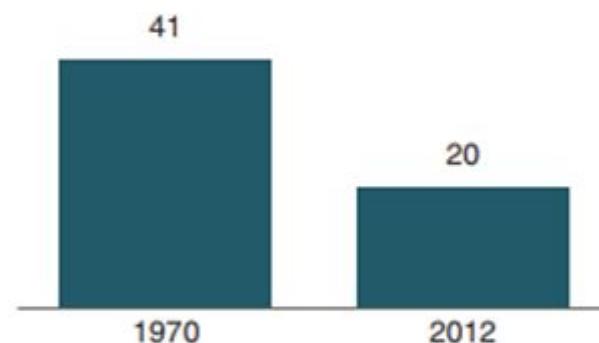


## Simple text



### 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

- Chỉ sử dụng số và mô tả ngắn gọn để làm nổi bật thông điệp chính

# Bảng



**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

- Nhiều trường thông tin cần so sánh và số lượng thuộc tính ở mức vừa phải, không quá nhiều

## Bản đồ nhiệt (Heatmap)

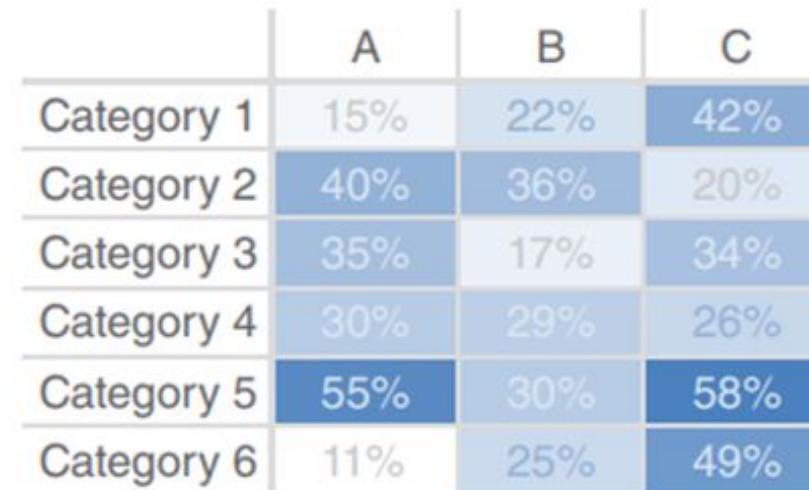


Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

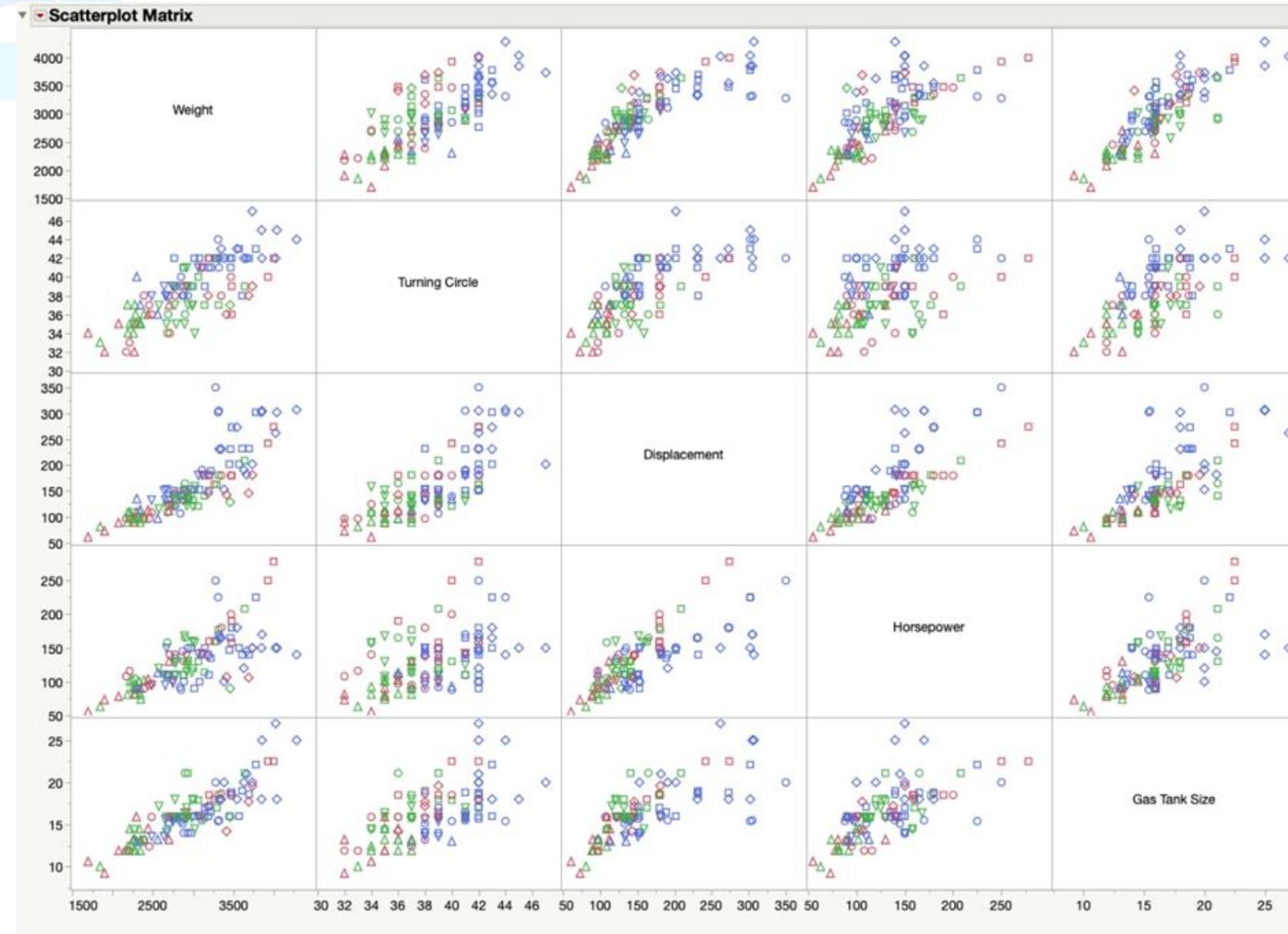
Heatmap

LOW-HIGH



- Dữ liệu dạng bảng
- Việc kết hợp màu sắc với trọng số sẽ giúp người đọc nhanh chóng nhận ra các miền giá trị trong bảng

## Scatter plot



- Mối quan hệ giữa hai đối tượng
- Mối quan hệ giữa biến phụ thuộc và các biến độc lập

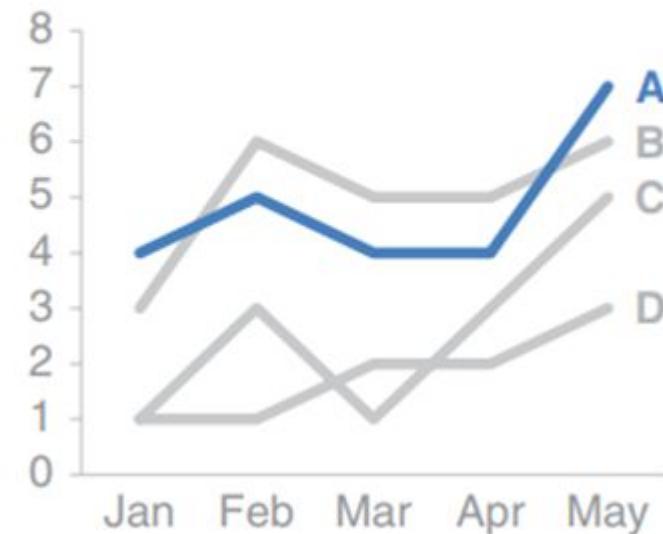


## Biểu đồ đường (Line graph)

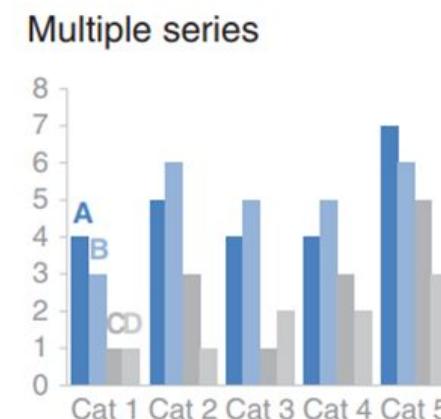
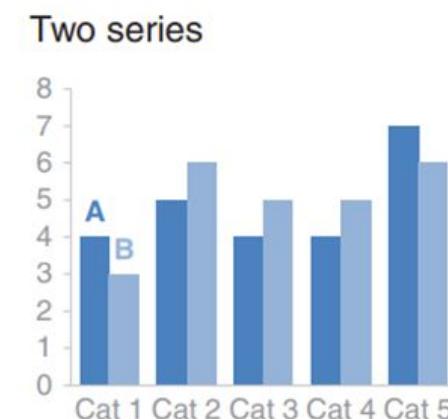
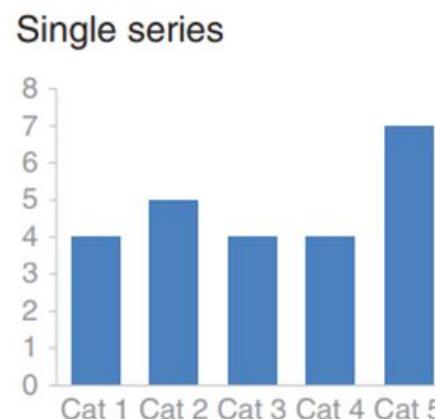
- Biểu diễn các **xu hướng**, những **thay đổi** theo thời gian và mối quan hệ giữa các biến liên tục
  - Theo dõi xu hướng theo thời gian: Doanh thu hàng tháng, tăng trưởng dân số hoặc thay đổi nhiệt độ trung bình năm.
  - So sánh nhiều chuỗi dữ liệu:** Khi có nhiều đối tượng khác nhau cần được so sánh trong cùng một khoảng thời gian, nhiều đường biến thiên trên cùng một đồ thị hiển thị sự thay đổi của từng đối tượng



Multiple series

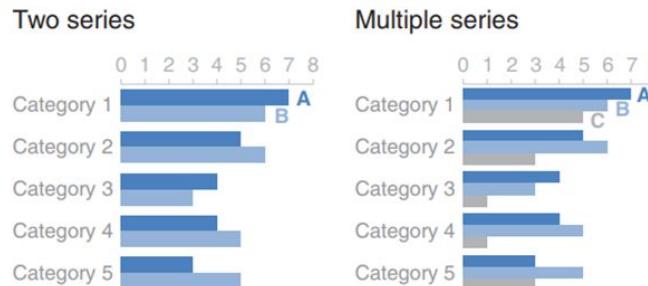
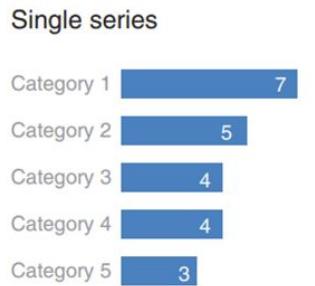
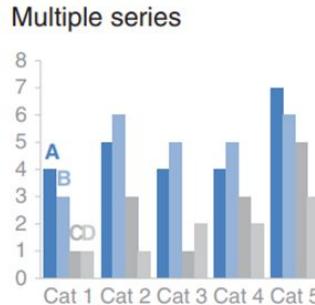
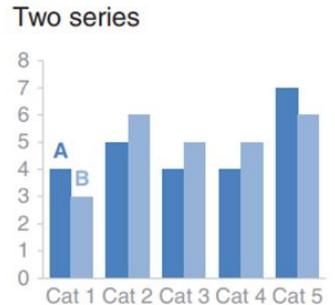
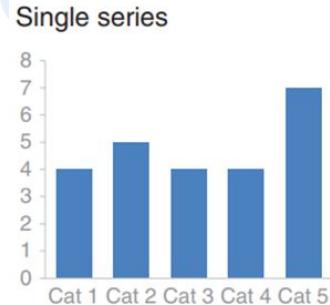


# Biểu đồ cột (Bar charts)



- Biểu diễn các biến phân loại
- Điểm nhìn tập trung vào ngọn của các cột, giúp người đọc nhanh chóng nhận diện các giá trị lớn nhất/ nhỏ nhất và khoảng cách giữa các đối tượng

# Horizontal bar chart/ Biểu đồ cột theo phương ngang



## Survey results

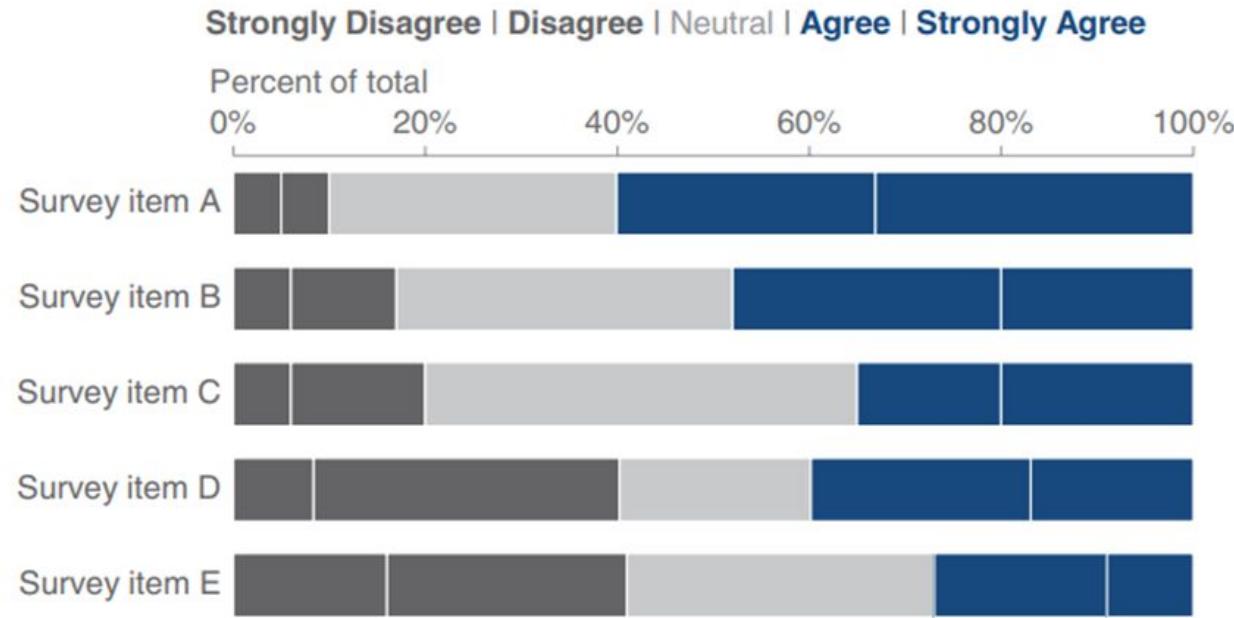


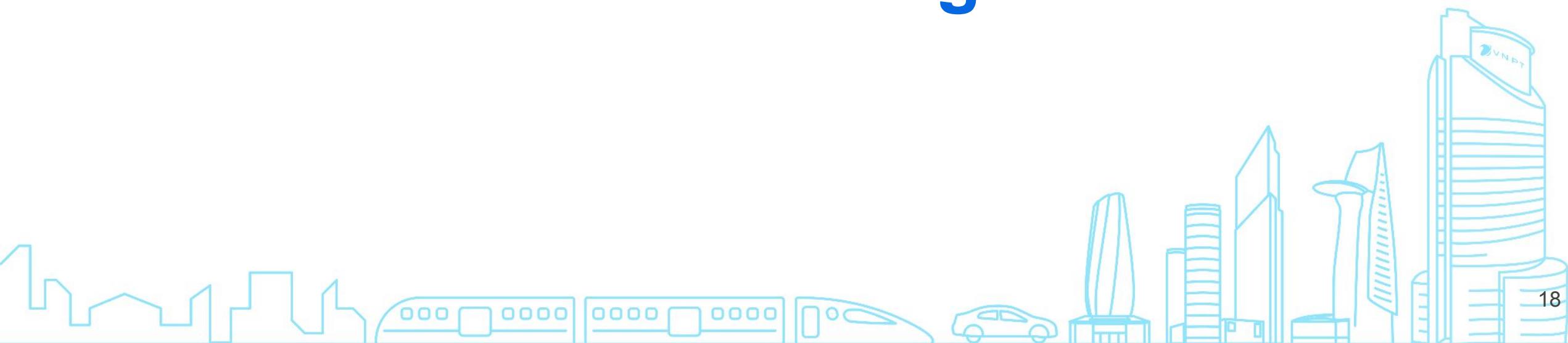
FIGURE 2.19 100% stacked horizontal bar chart

- Tên đối tượng dài khó biểu diễn
- Xếp các đối tượng cần quan tâm lên dòng đầu tiên vì đó là vị trí dễ nhìn nhất
- Biểu đồ cột chồng (Tự đọc)



3

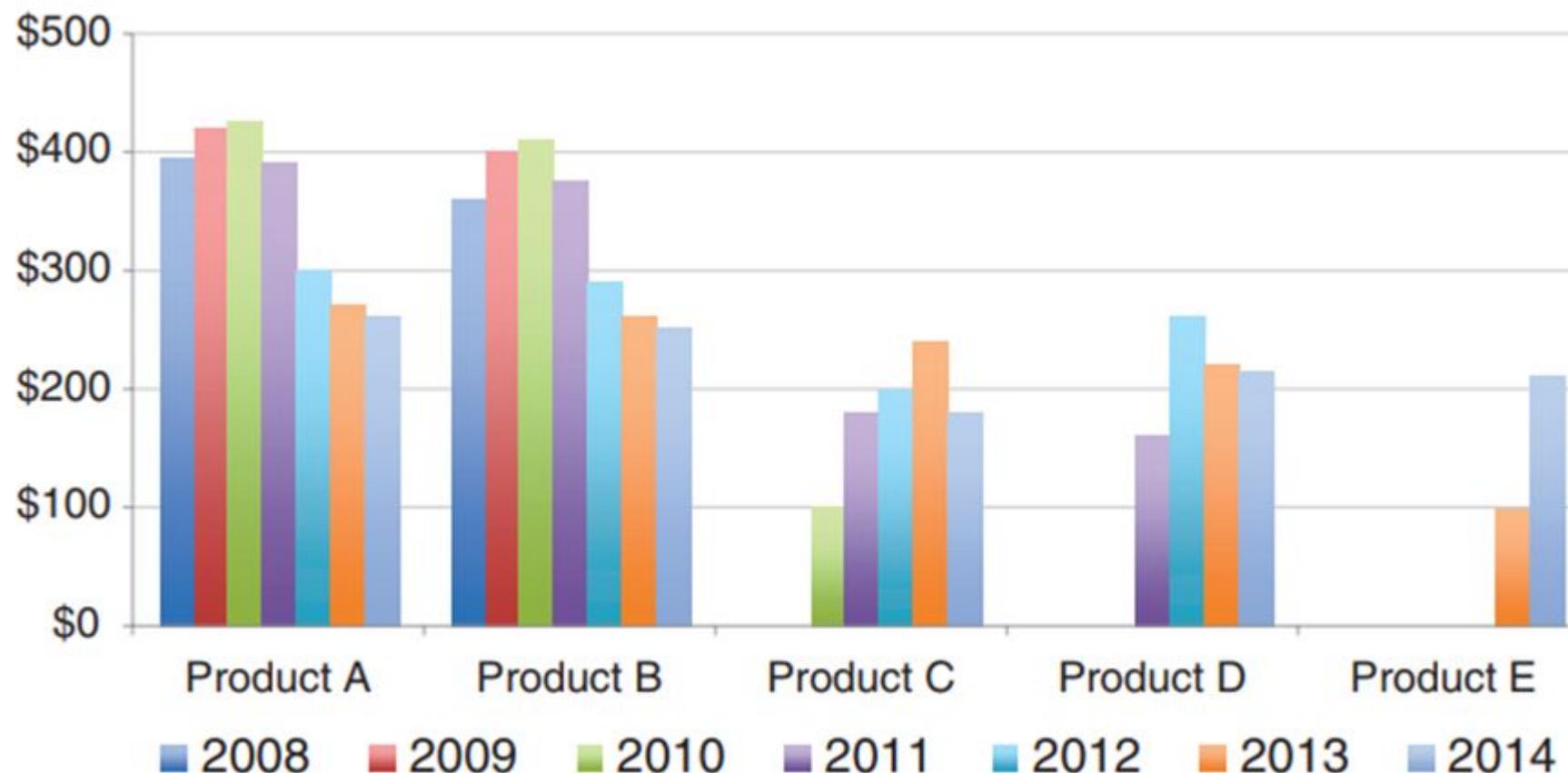
## Decluttering





# Why Decluttering?

Average Retail Product Price per Year

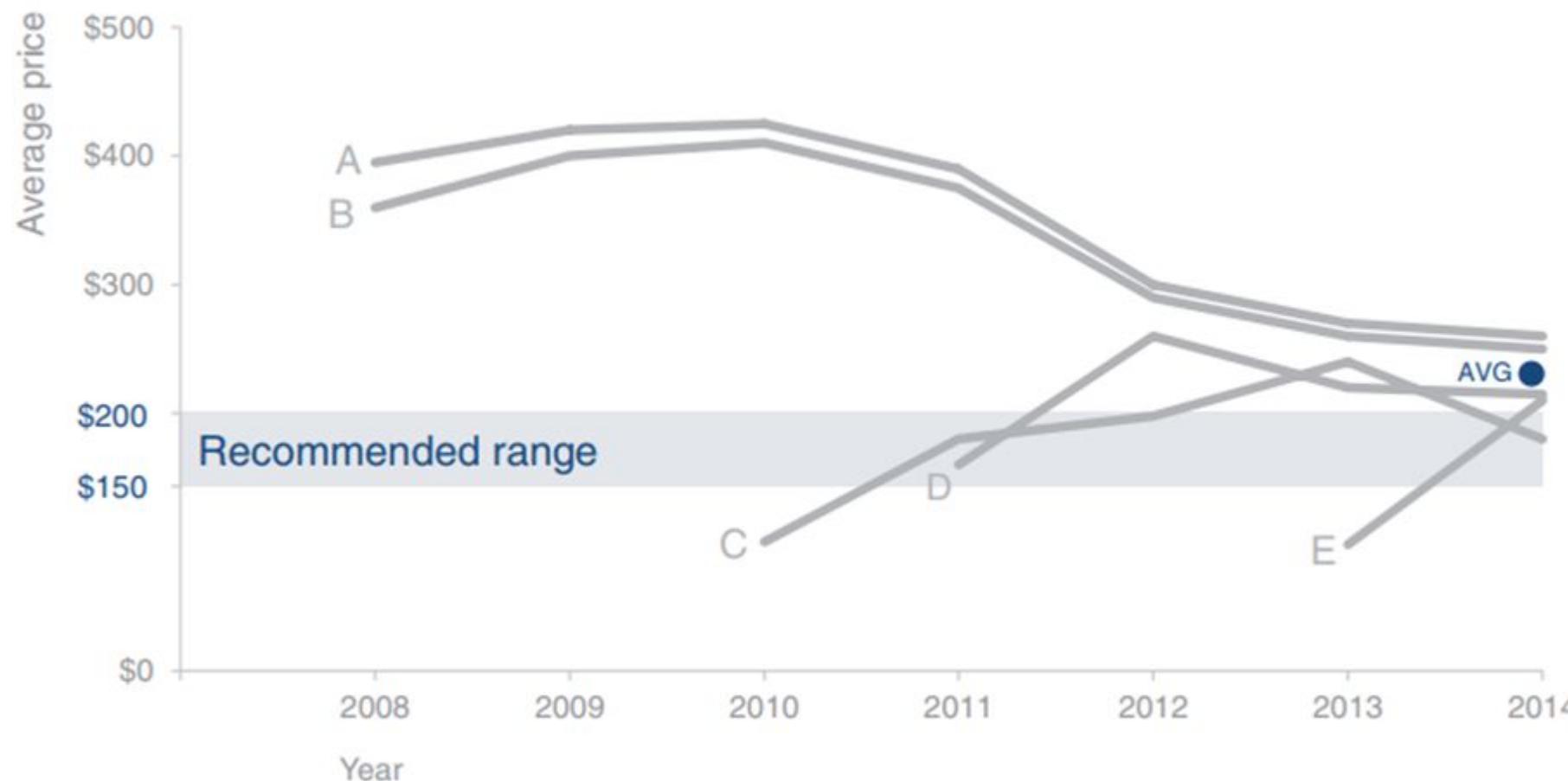




# Why Decluttering?



Retail price over time by product

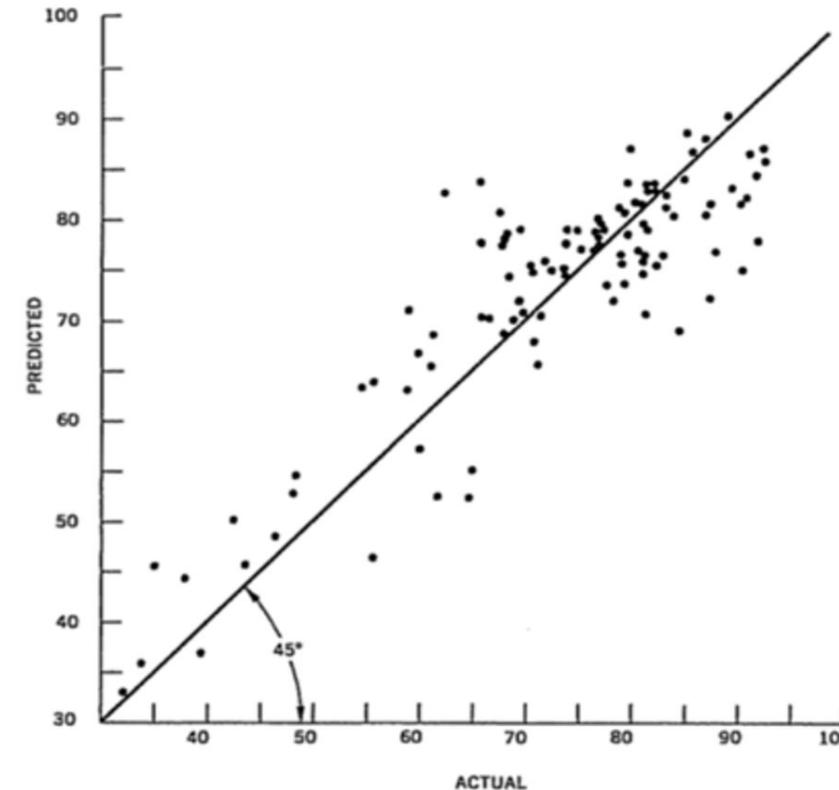
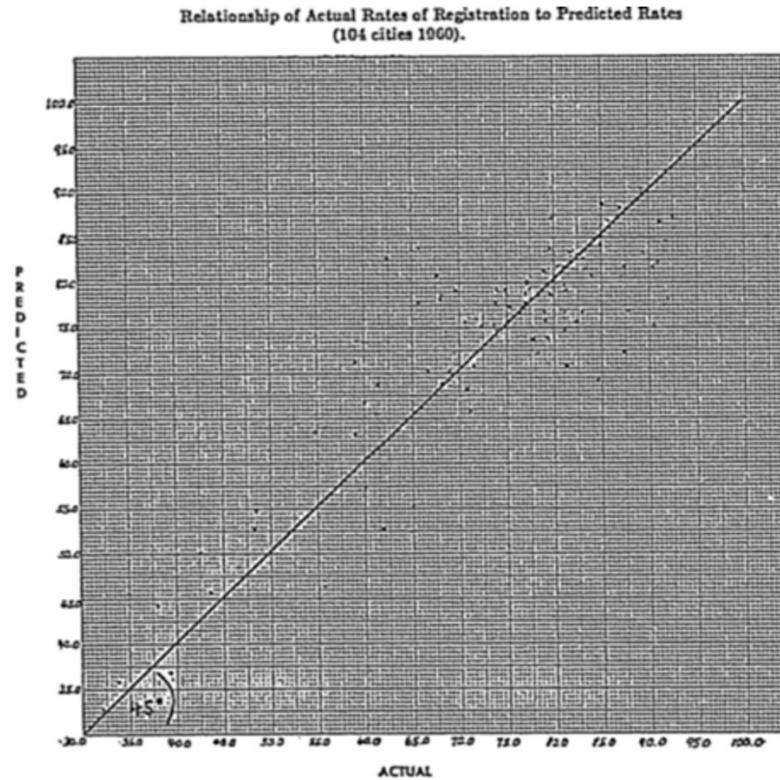


## Cognitive Load (Khả năng nhận thức)

- Thông tin từ 2 biểu đồ giống hay khác nhau?
- Biểu đồ nào giúp bạn tìm được Insights nhanh hơn, dễ dàng hơn?
- Sức mạnh não bộ -> Mức độ/ Khả năng nhận thức có giới hạn
  - => Thúc đẩy quá trình nhận thức này nhanh chóng và dễ dàng là chìa khoá giúp người xem, người nghe dễ nắm bắt thông tin hơn và tạo điều kiện thuận lợi cho quá trình ra quyết định
  - => Loại bỏ nhiễu và Làm nổi bật thông tin chính



## Data-ink ratio



- Data-ink ratio = Lượng mực sử dụng để biểu diễn dữ liệu/ Lượng mực sử dụng trong toàn bộ biểu đồ
- Lượng mực để biểu diễn dữ liệu nhiều hơn so với các chi tiết không phải dữ liệu thì càng tốt  
=> Hạn chế chi tiết thừa

# Gestalt Principles of Visual Perception

## *Gestalt Principles of Visual Perception*

Principle	Description
Proximity	Objects that are close together are perceived as a group.
Similarity	Objects that share similar attributes (e.g., color or shape) are perceived as a group.
Enclosure	Objects that appear to have a boundary around them (e.g., formed by a border or area of common color) are perceived as a group.
Closure	Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.
Continuity	Objects that are aligned together or appear to be a continuation of one another are perceived as a group.
Connection	Objects that are connected (e.g., by a line) are perceived as a group.

**Proximity:** **A B C D E**  
**F G I J K L**



**Similarity:** **A B C D E**  
**F G I J K L**



**Continuity:** **A B C D E K L**  
**F G**



**Closure:** **A B C D E**  
**F G I J K L**



**Connection:** **A B C D E**  
**F G I J K L**

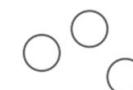


**Figure/Ground:** **A B C D E**  
**F G I J K L**

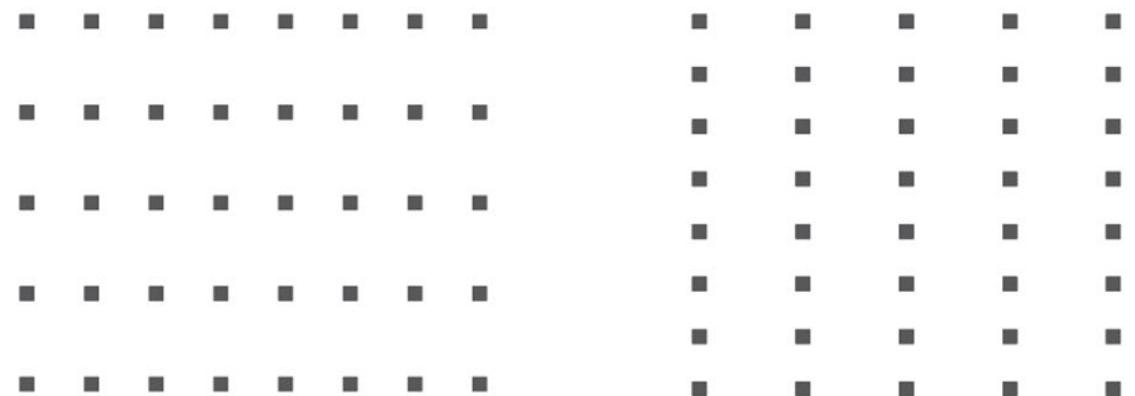
# Gestalt Principles of Visual Perception

## Proximity - Tính tiệm cận

- Chúng ta nhận thức các vật thể gần nhau sẽ cùng thuộc về một nhóm

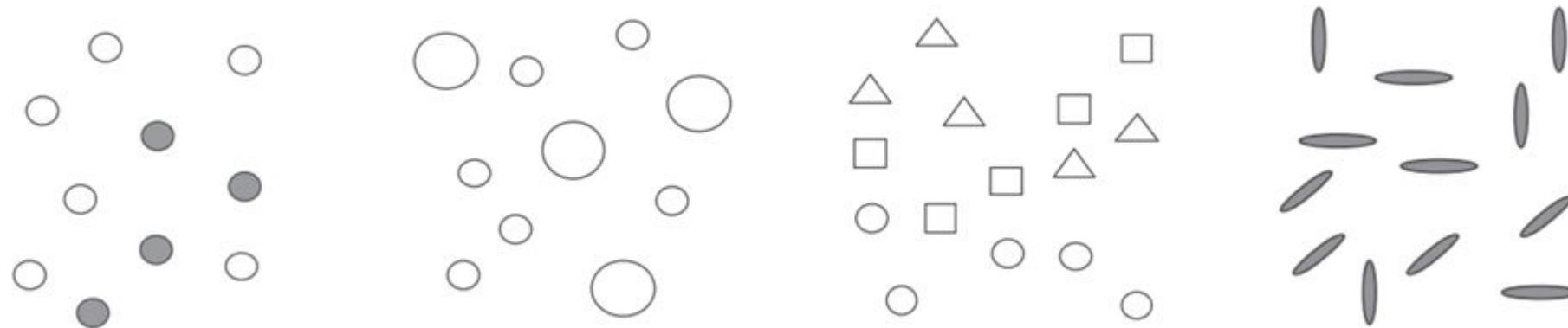


- Ứng dụng: thiết kế bảng biểu định hướng người đọc đọc theo hàng ngang hay cột dọc



# Gestalt Principles of Visual Perception

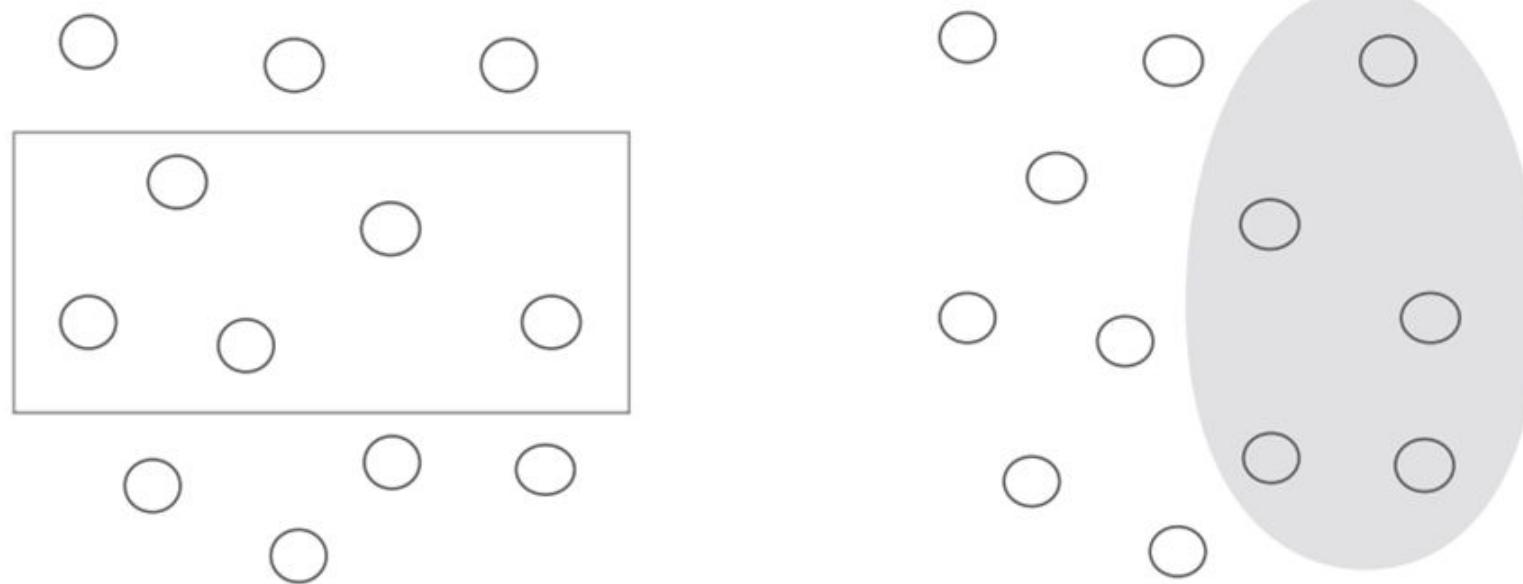
## Similarity – Tính tương tự



- Các vật thể tương tự nhau về màu sắc, kích cỡ, hình dạng hoặc phương hướng thường được nhận thức là cùng thuộc về một nhóm
- Ứng dụng: mã hoá và đại diện cho các dữ liệu có đặc tính giống nhau hoặc muôn gộp thành nhóm khi vẽ biểu đồ
- Lưu ý: quá nhiều đặc tính -> quá nhiều màu sắc, hình dáng, kích thước, etc. -> nhiễu

## Gestalt Principles of Visual Perception

### Enclosure – Tính vây quanh



- Chúng ta nhận diện các vật thể trong khung vuông và hình tròn cùng thuộc về một nhóm do có các đường bao, đường viền giới hạn xung quanh chúng
- Ứng dụng thường xuyên trong các bảng và biểu đồ bằng cách sử dụng đường viền và màu nền để nhóm thông tin và đôi khi để nhấn mạnh các đối tượng cụ thể

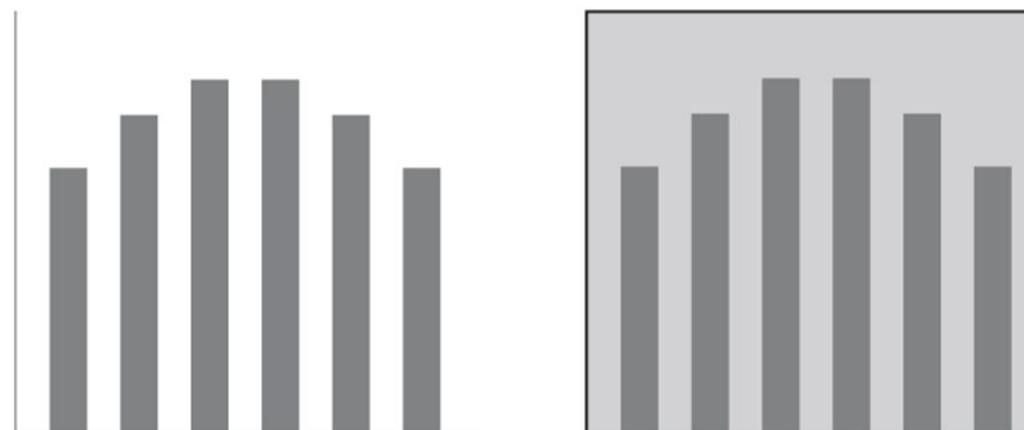


## Gestalt Principles of Visual Perception

### Closure – Tính đóng/ tính hoàn thiện



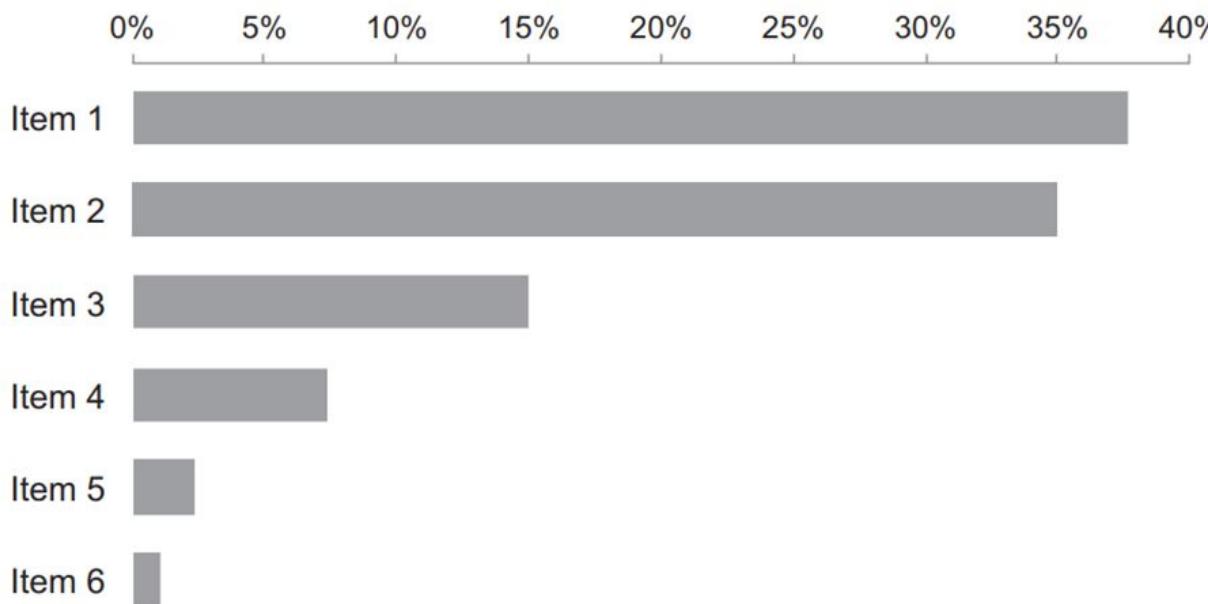
- Chúng ta có xu hướng nhận thức các vật thể có tính toàn diện và nhìn sự vật dưới con mắt giải thích cho những yếu tố mơ hồ, chưa hoàn thiện



# Gestalt Principles of Visual Perception

## Continuity – Tính liên tục

- Chúng ta coi các vật thể là thuộc cùng một nhóm, là một phần của một tổng thể duy nhất, nếu chúng thẳng hàng với nhau hoặc liên tục, nối tiếp nhau.

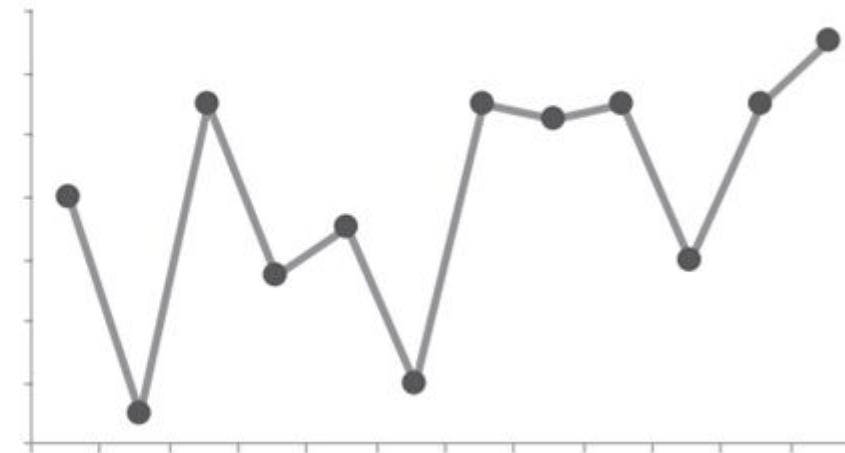
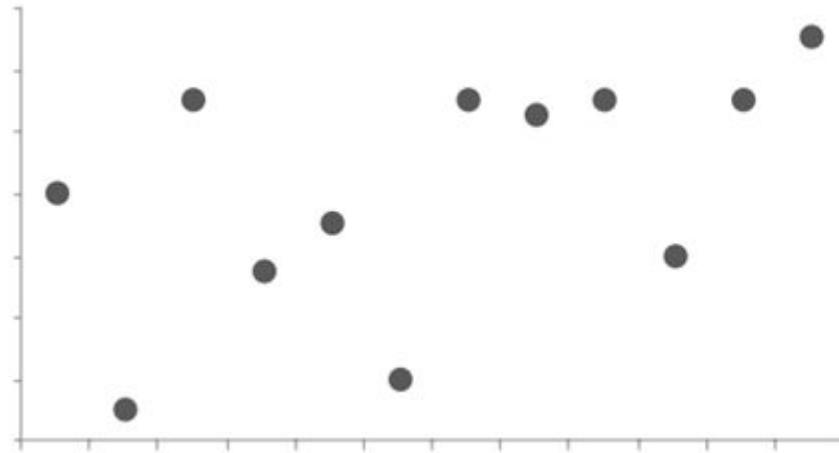


Division/Department	Headcount
G&A	
Finance	15
Purchasing	5
Information Systems	17
Sales	
Field Sales	47
Sales Operations	10
Engineering	
Product Development	22
Product Marketing	5



## Gestalt Principles of Visual Perception

### Connection – Tính liên kết



- Chúng ta nhận diện các vật thể được nối với nhau bởi đường nối sẽ có điểm chung và cùng thuộc về một nhóm
- Ứng dụng: Dữ liệu chuỗi thời gian, Network science

## Tufte's 6 Principles of Graphical Integrity

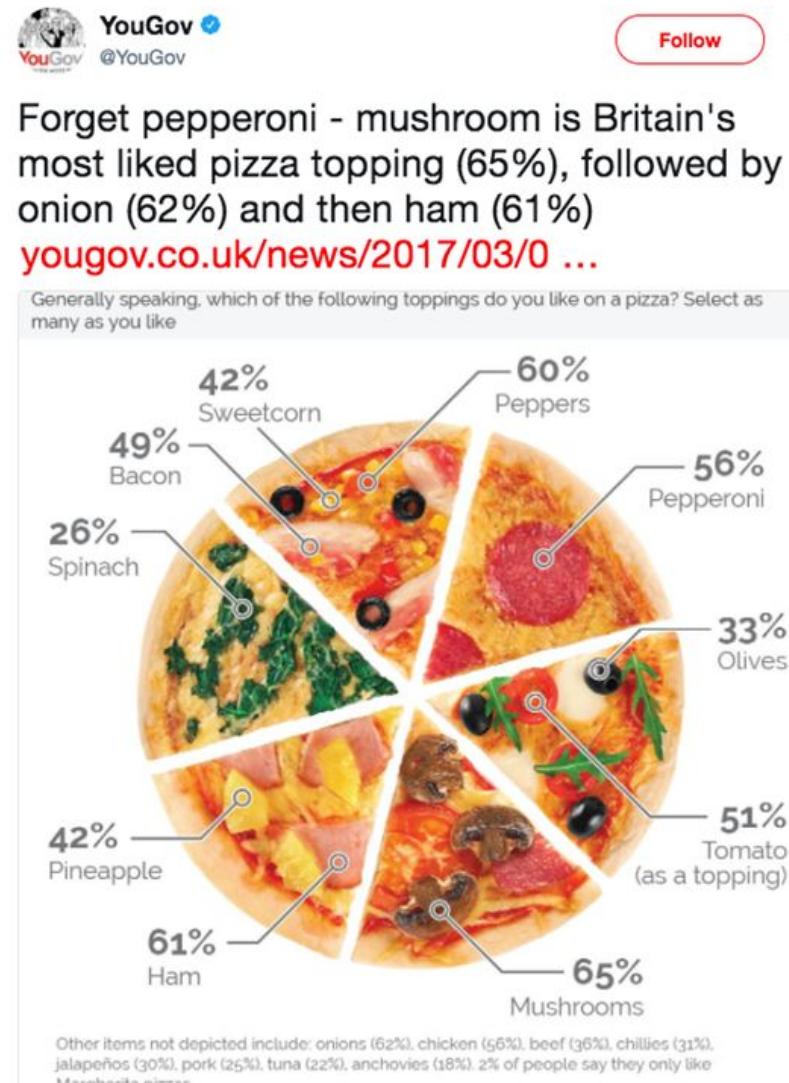
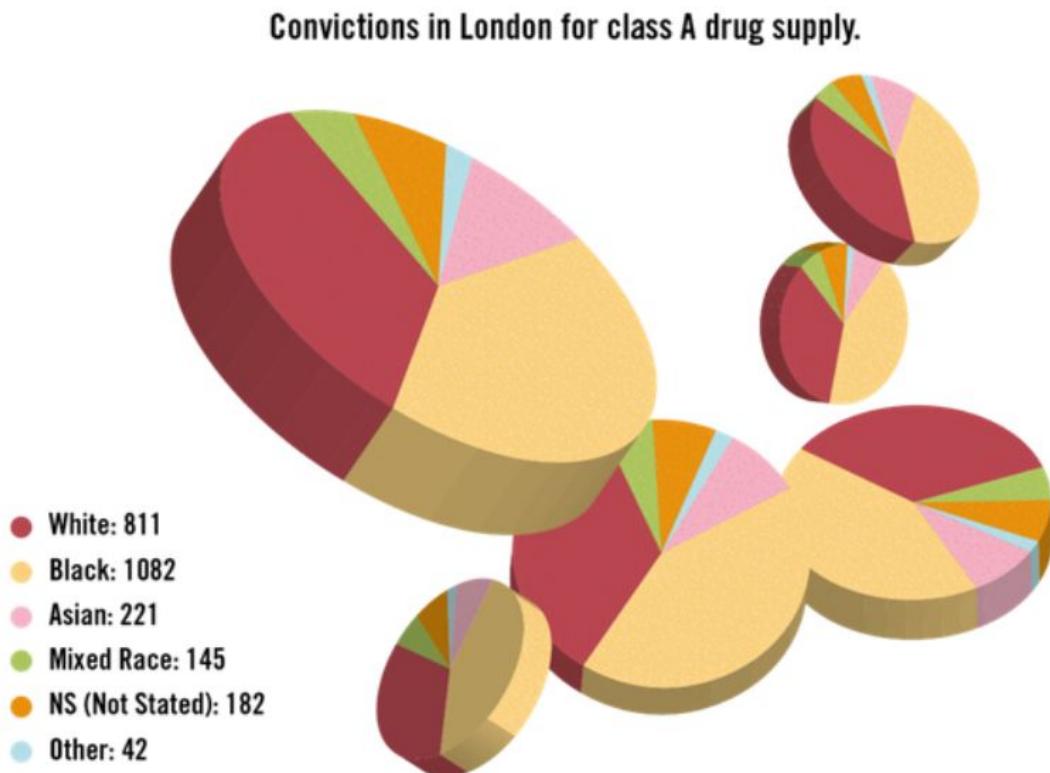
1. **Proportionality:** “The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.”
2. **Clarity:** “Clear, detailed, and thorough labelling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.”
3. **Data driven:** “Show variation that is in the data, not design variation.”
4. **Standardised:** “In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.”
5. **Appropriate Dimensions:** “ The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”
6. **Contextualised:** “Graphics must not quote data out of context.”



“Good design is clear thinking made visible”



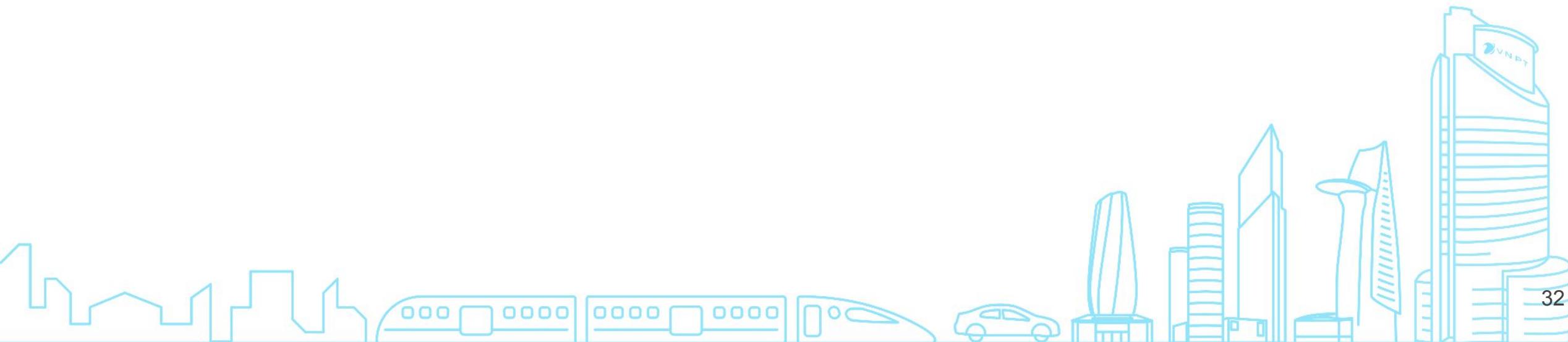
# Một số biểu đồ vi phạm nguyên tắc Tuft





4

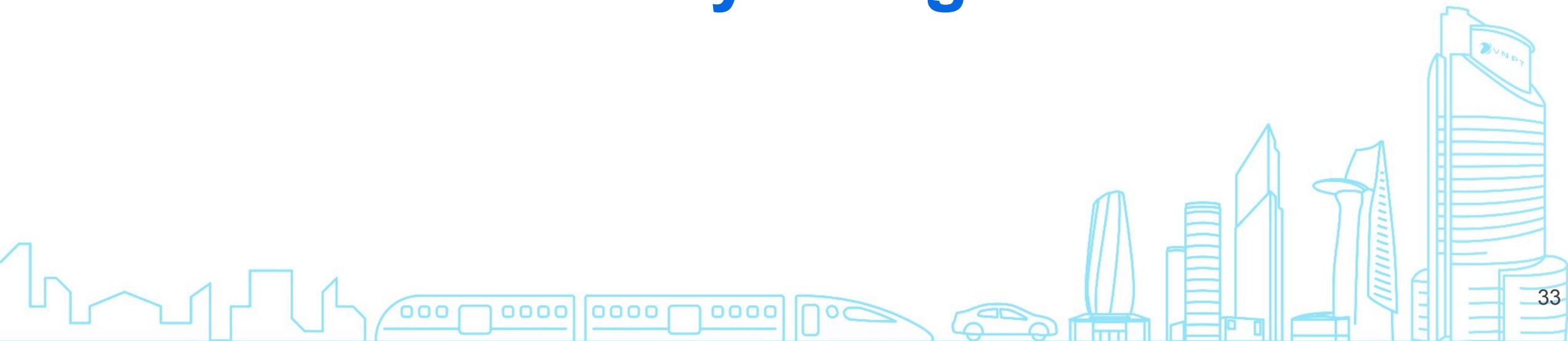
## Làm nổi bật dữ liệu và hoàn thiện biểu đồ





5

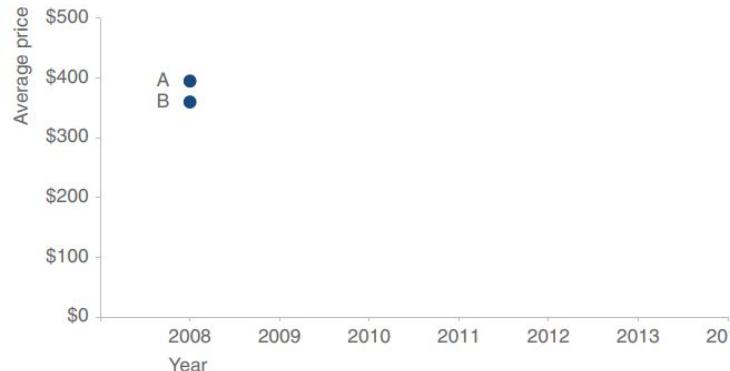
# Storytelling



# Storytelling

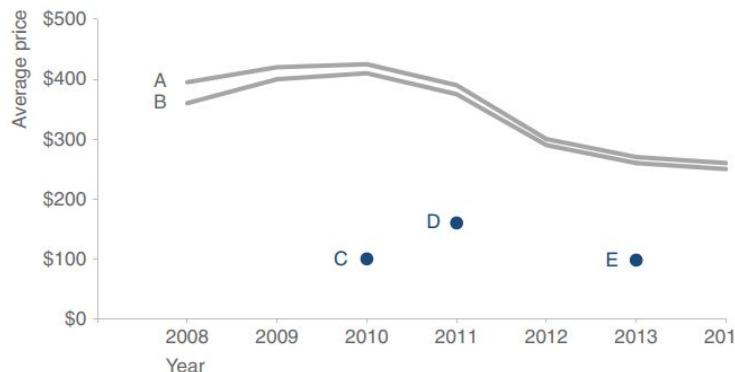
Products A and B were launched in 2008 at price points of **\$360+**

Retail price over time



Products C, D, and E were each introduced later at **much lower price points...**

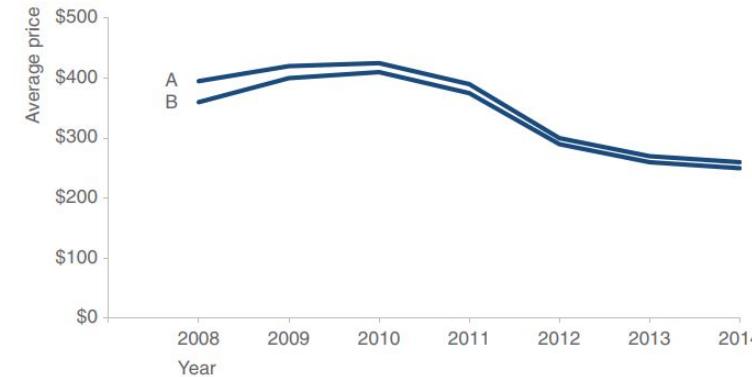
Retail price over time



# Storyboard

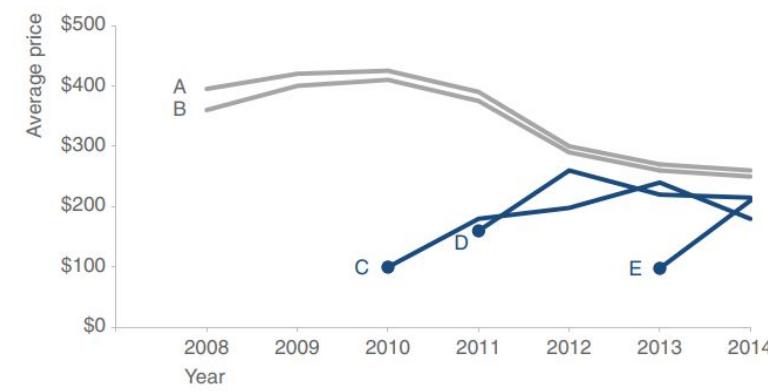
They have been priced similarly over time, with B consistently slightly lower than A

Retail price over time



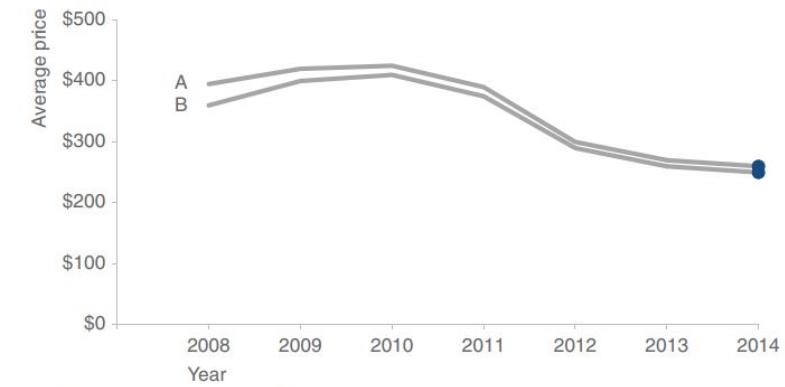
...but all have **increased in price** since their respective launches

Retail price over time



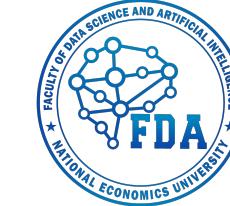
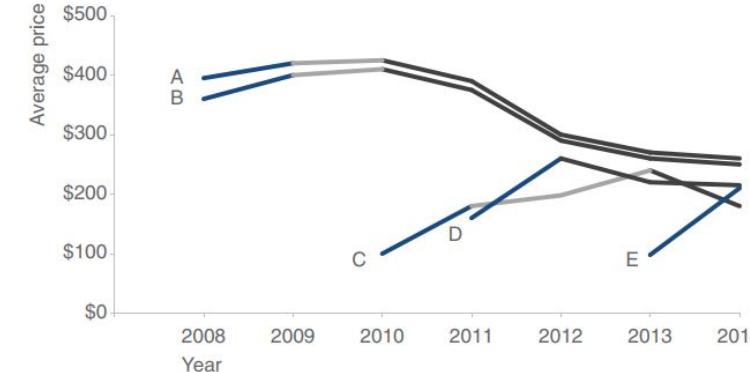
In 2014, Products A and B were priced at **\$260** and **\$250**, respectively

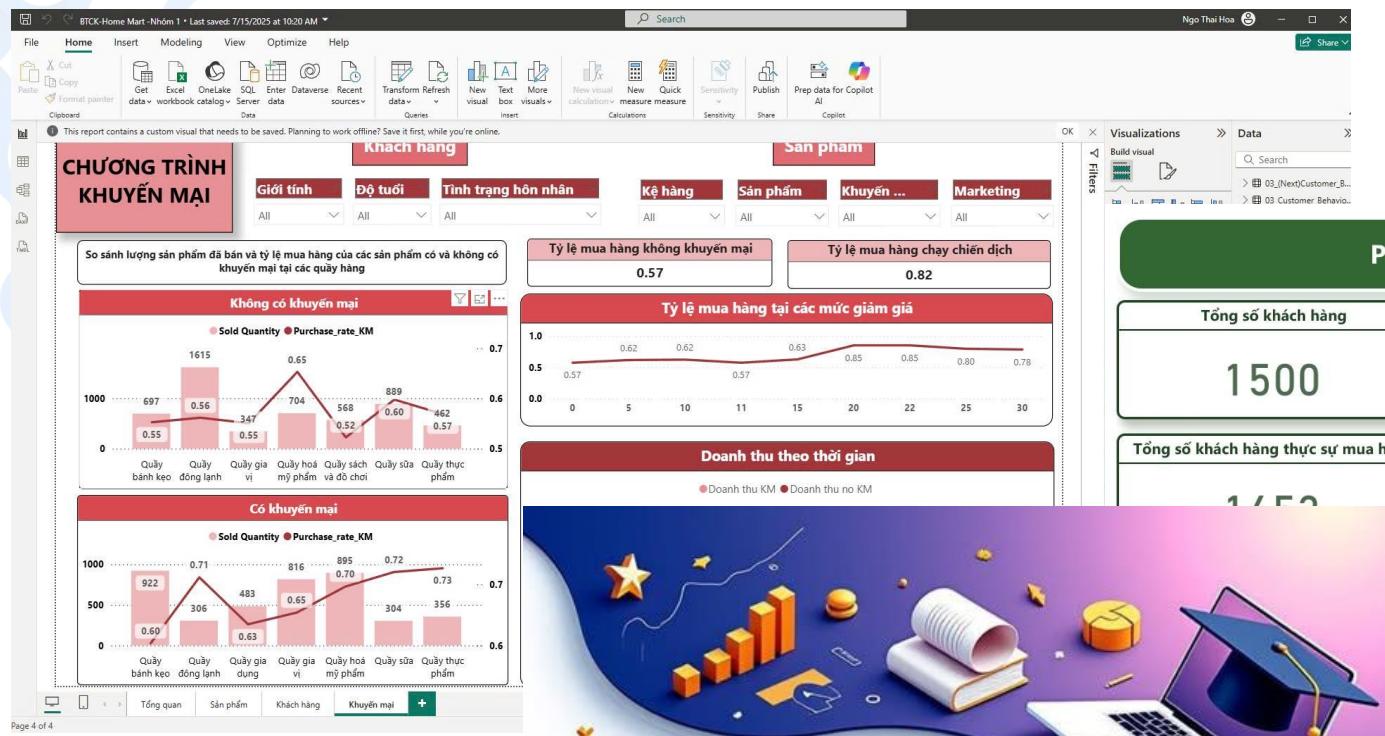
Retail price over time



In fact, with the launch of a new product in this space, we tend to see an **initial price increase**, followed by a **decrease over time**

Retail price over time



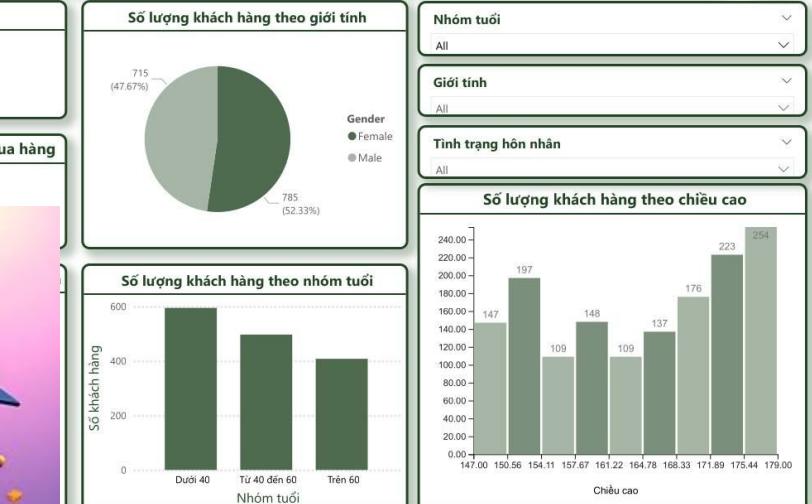


Page 4 of 4



Hotline: +84 366902278 | Fanpage khóa học: facebook.com/neu.vnpt/ | Website: fda.neu.edu.vn/ | Địa chỉ: P.1604, Tòa A1, Trường Công nghệ | ĐĂNG KÝ NGAY | <https://www.facebook.com/neu.vnpt>

## PHÂN TÍCH ĐẶC ĐIỂM KHÁCH HÀNG



**Khóa Học Phân Tích Dữ Liệu Kinh Doanh  
NEU - VNPT AI**



# Homework

- Các nhóm xây dựng 2 biểu đồ đã được phân công
- Áp dụng quy trình biểu diễn dữ liệu và Chỉ ra cách áp dụng
- Nộp bài trên Ims
- Code và trình bày kết quả trên lớp





# Effective Data Visualisation & Storytelling

Khoa Khoa học Dữ liệu và Trí tuệ nhân tạo (FDA)  
Trường Công nghệ, Trường Đại học Kinh tế Quốc dân

VNPT AI  
Tập đoàn Bưu chính Viễn thông Việt Nam VNPT

# Thank you!

