

# Quiz 1 - Solutions and discussion

## ■ Autores

◆ João Moura Pires ([jmp@fct.unl.pt](mailto:jmp@fct.unl.pt))

- This material can be freely used for personal or academic purposes without any previous authorization from the author, provided that this notice is maintained/kept.
- For commercial purposes the use of any part of this material requires the previous authorization from the author.

# Motivation and Goals of Data Visualization, Perception and Principles

Choose the sentence that better describes the **ultimate goal** of Data Visualization:

Select one:

- ☐ To build, from data, synthetic images that will be appreciated by the general public.
- ☐ To efficiently use computer graphics techniques and algorithms to build computer images.
- ☐ To build graphical user interfaces to query data files.
- ☒ To understand the data to take decisions.

The purpose of data visualization is to help the user to understand the data to make decisions. What is the central idea of how data visualization is performed? Choose the correct option.

Select one:

- ☒ A mapping from data variables to visual variables to build an image.
- ☐ By representing categorical data variables using color and the numerical variables using scales.
- ☐ By dividing the screen in cells and color them to build an image.
- ☐ A mapping from visual variables to data variables build an image.
- ☐ By dividing the data in small subsets that can fit the screen resolution.



# The role of Interactivity in Data Visualization Process

The purpose of data visualization is to help the user to understand the data in order to make decisions. What is the role of interactivity in this process? Select the correct sentence.

Select one:

- ☐ To change the data variables that will be mapped to the screen.
- ☐ To play and control the visual animations.
- ☒ To enable the user select different subsets of data and different mapping to the visual variables.
- ☐ To filter the data present in the visualization.
- ☐ To highlight in the visualization the desired data.

# Lesson learned from Anscombe dataset

Anscombe proposed 4 datasets, with 2 variables X and Y, represented in the figure below, that share the values of a few statistics like the mean of X, the mean of Y, the variance of X, the variance of Y, and the correlation between X and Y.

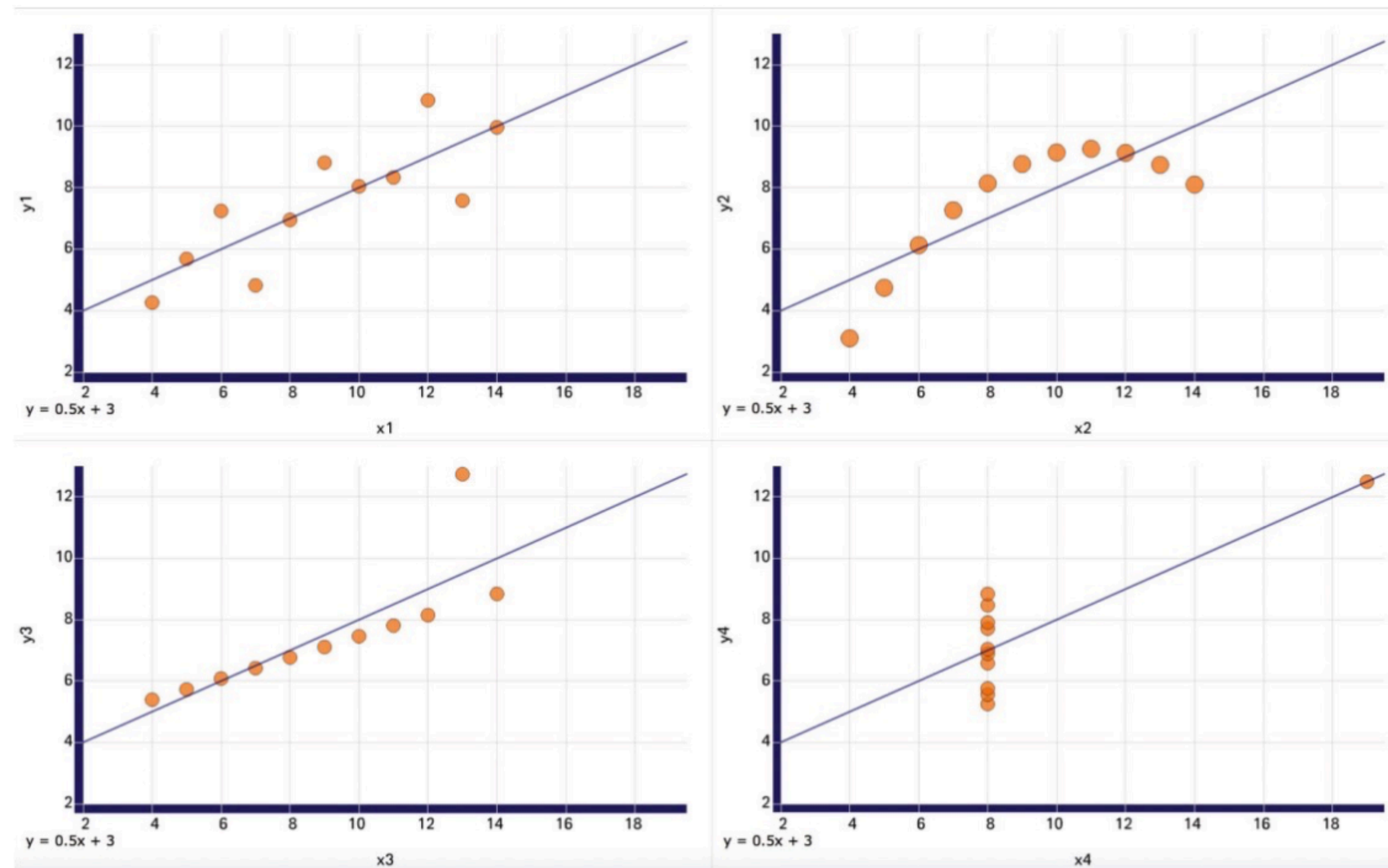


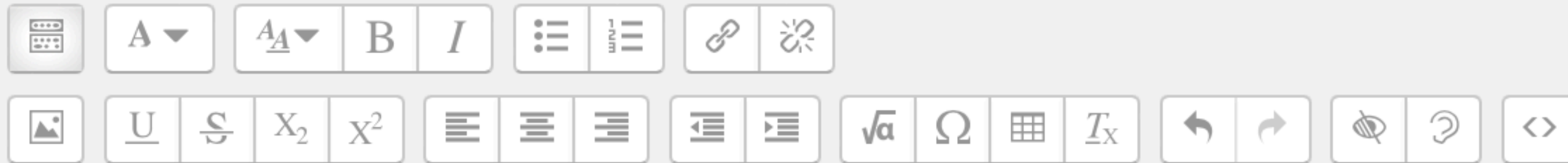
Figure 1.22. Plots of four data sets with identical statistics.

Write a **short** sentence describing what is the **lesson learned** from this dataset:

Answer:

The well-known perceptual color models, initiated by Munsel, has three essential characteristics: a) Perceptual balance; b) Distinguishability; c) Flexibility.

Explain in your own words what each of these features means.



- **Perceptual balance.** A unit step anywhere along the color scale produces a perceptually uniform difference in color.
- **Distinguishability.** Within a discrete collection of colors, every color is equally distinguishable from all the others (i.e., no specific color is “easier” or “harder” to identify).
- **Flexibility.** Colors can be selected from any part of color space (e.g., the selection technique is not restricted to only greens, or only reds and blues).



What are pre-attentive properties?

Select one:

- ☐ A significant set of visual properties that are detected very quickly (between 200 and 250 ms) and with great precision by the low-level vision system.
- ☐ A limited set of visual properties that are detected by the low-level vision system before we pay attention.
- ☒ A limited set of visual properties that are detected very quickly (between 200 and 250 ms) and with great precision by the low-level vision system.
- ☐ A limited set of visual properties that are detected very quickly (between 200 and 250 ms) but not very accurately by the low-level vision system.

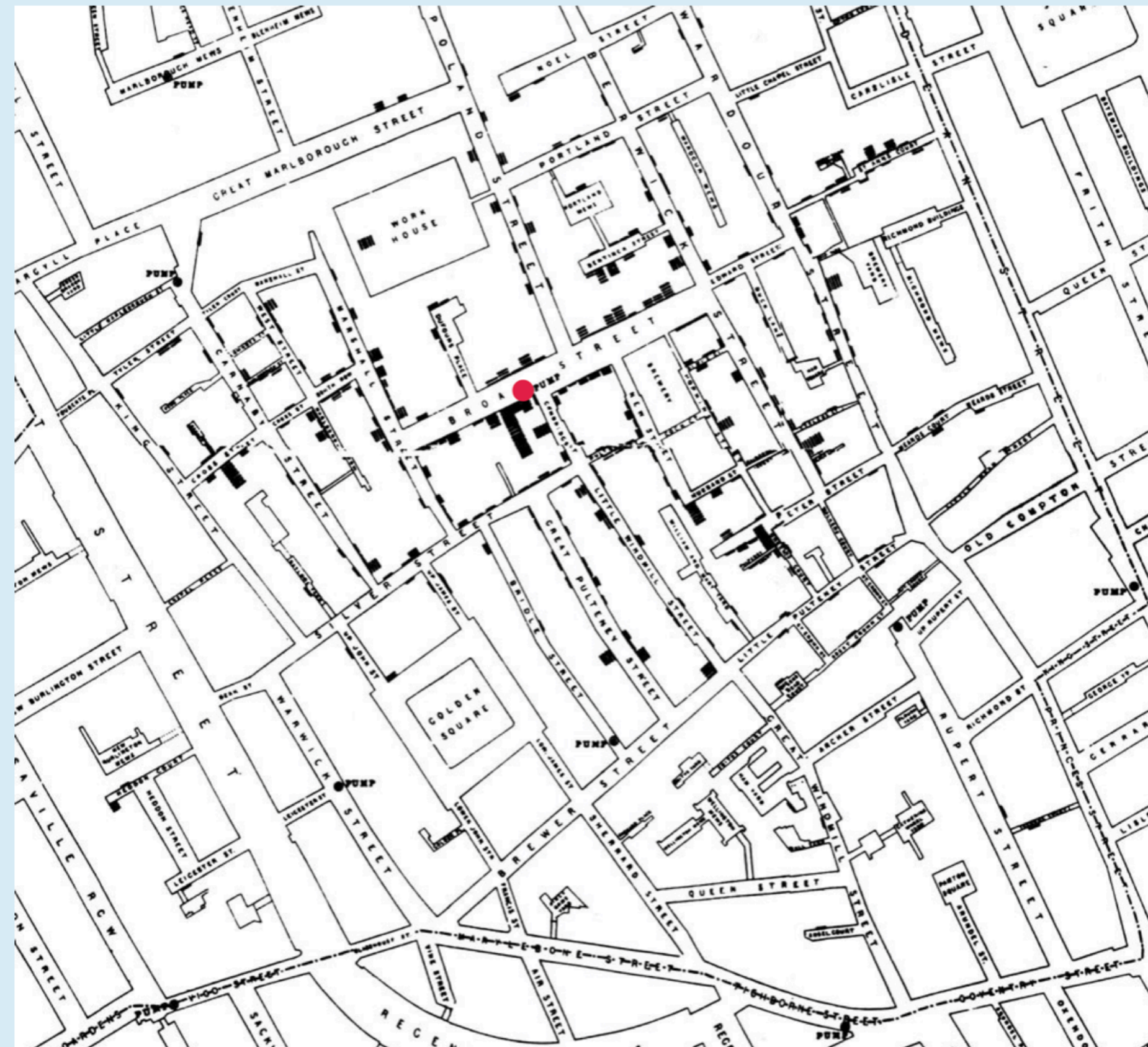
One of the most famous data visualizations is the Map of John Snow that represents the Broad Street cholera outbreak in London in 1854. Provide a short answer for:

(a) indicate, for each variable of the data present on the map and what is its mapping to visual variables:

Death address - location; Number of deaths - size of bars

(b) indicate the main discovery made by John Snow: The disease is being propagated through the water pump marked i

(c) and the main decision taken from there: Close the Pump





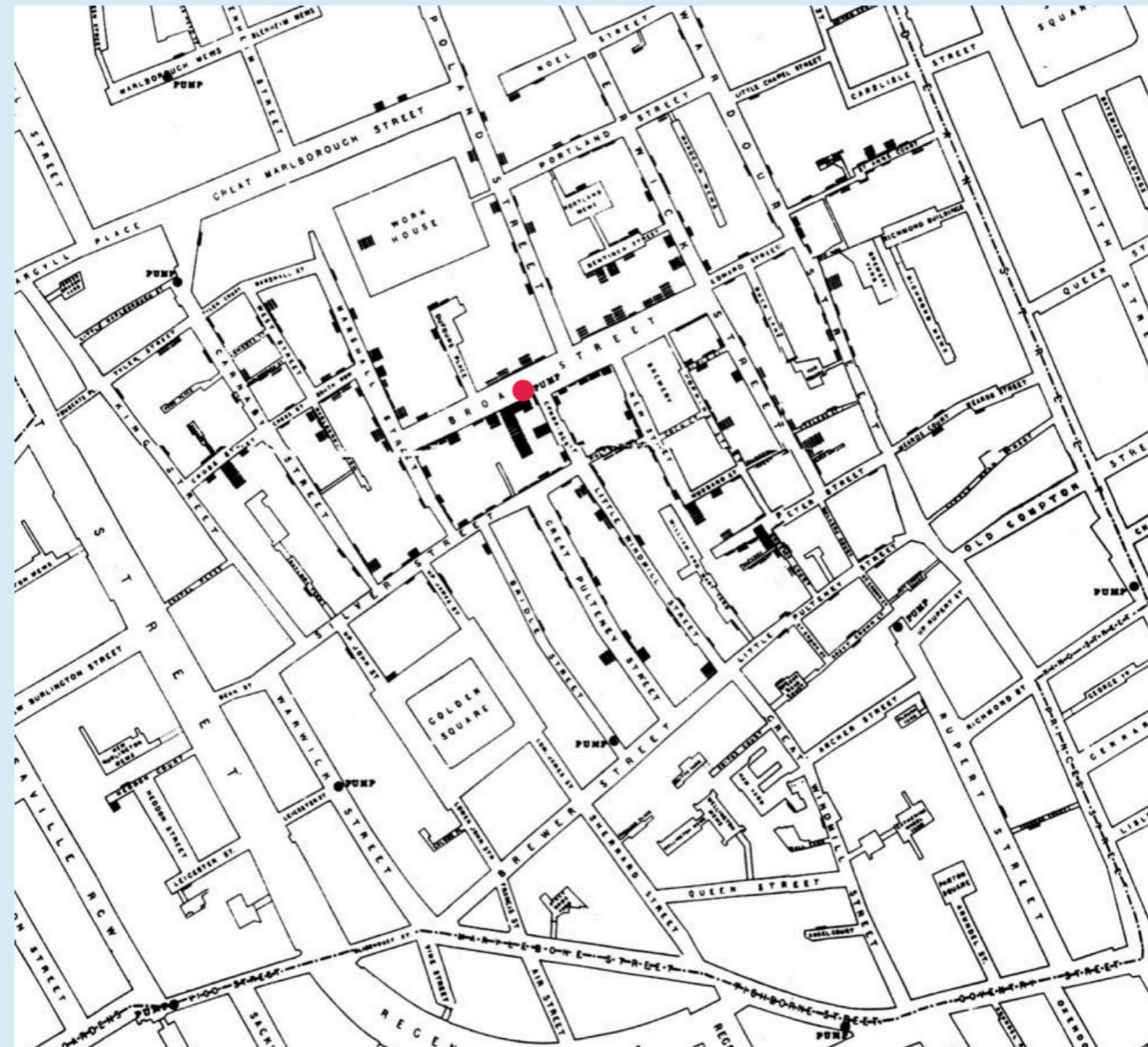
One of the most famous data visualizations is the Map of John Snow that represents the Broad Street cholera outbreak in London in 1854. Provide a short answer for:

(a) indicate, for each variable of the data present on the map and what is its mapping to visual variables:

Death address - location; Number of deaths - size of bars

(b) indicate the main discovery made by John Snow: The disease is being propagated through the water pump marked i

(c) and the main decision taken from there: Close the Pump



## Basic Data Visualization (may include a Tableau)



There are 4 types of measurement scale: nominal, ordinal, interval, and ratio. For each of the data variables, indicate which type of measurement scale is most suitable

Full Name	Nominal
Quantity	Ratio
Product Category	Nominal
Temperature (celcius)	Interval
Sex	Nominal
Price	Ratio

Select the CORRECT sentence. Dimensionality reduction is a technique used:

Select one:

- ☐ a. In data visualization, to address the issues of missing data in a multivariate dataset.
- ☒ b. In data visualization, to reduce the number of columns in a multivariate dataset.
- ☐ c. In data visualization, to reduce both the number of columns and the number of rows in a multivariate dataset.
- ☐ d. In data visualization, to reduce the number of rows in a multivariate dataset.

For each variable from Cars Dataset used in the lab sessions, indicate the most appropriate data type, in terms of data visualization: A: Nominal; B: Categorical; C: Ordinal; D: Numeric - Discrete; E: Numeric - Continuous.

Price Class (cheap, medium, high)	C: Ordinal
N° Cylinders	D: Numeric - Discrete
CityMPG	E: Numeric - Continuous
Brand	B: Categorical
Quarter (Q1, Q2, Q3, Q4)	C: Ordinal
Model Name	A: Nominal
Car class	B: Categorical

For each variable from the sample Superstore workbook, indicate the most appropriate data type, in terms of data visualization:  
A: Nominal; B: Categorical; C: Ordinal; D: Numeric - Discrete; E: Numeric - Continuous.

Profit

E: Numeric - Continuous

Customer Name

A: Nominal

Quantity (integer)

D: Numeric - Discrete

Ship Status

B: Categorical

Quarter (Q1, Q2, Q3, Q4)

C: Ordinal

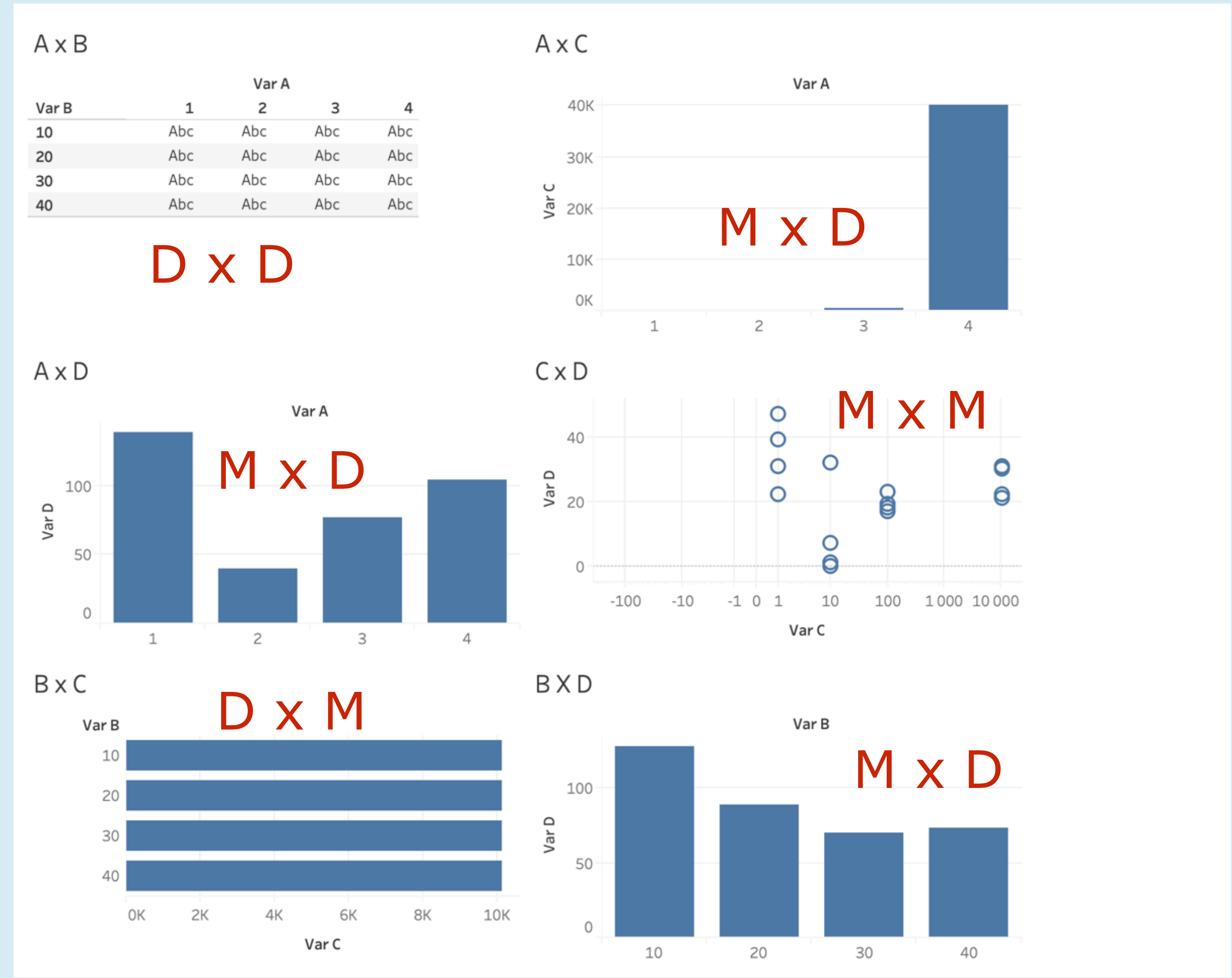
Customer Segment

B: Categorical



# Dimensions and Measures (A...D)

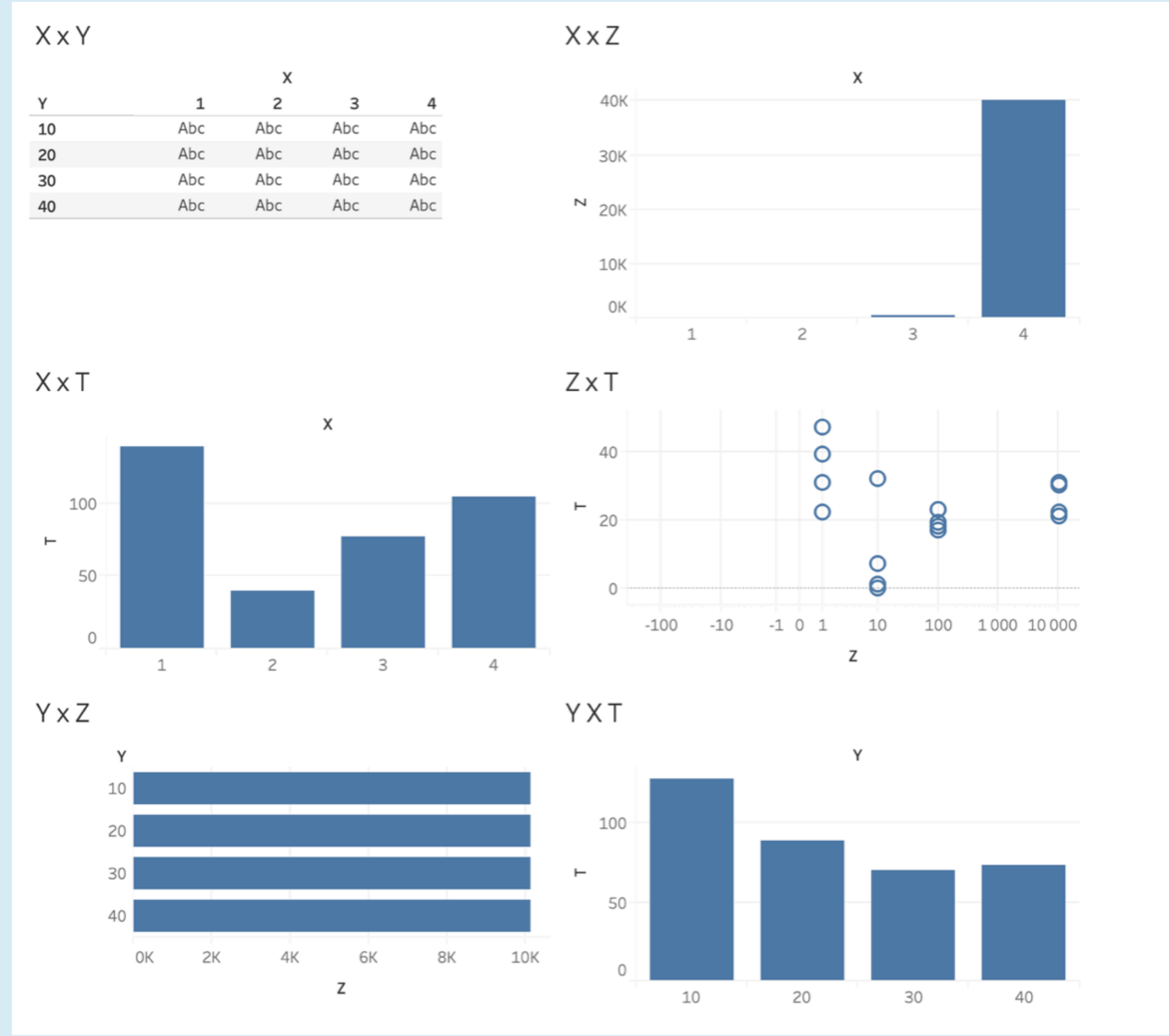
Consider the following graphics from a dataset with 4 variables named, varA, varB, varC e varD. Complete the next table with the letters M and D (M=Measure, D=Dimension) to indicate the role of each variable present in rows or columns for each graphic.



C x D	Rows: M; Columns: M
A x D	Rows: M; Columns: D
A x C	Rows: M; Columns: D
B x C	Rows: D; Columns: M
B x D	Rows: M; Columns: D
A x B	Rows: D; Columns: D

# Dimensions and Measures (X...Z)

Consider the following graphics from a dataset with 4 variables named, X, Y, Z e T. Complete the next table with the letters M and D (M=Measure, D=Dimension) to indicate the role of each variable present in rows or columns for each graphic.



Z x T	Rows: M; Columns: M
X x Z	Rows: M; Columns: D
X x T	Rows: M; Columns: D
Y x T	Rows: M; Columns: D
X x Y	Rows: D; Columns: D
Y x Z	Rows: D; Columns: M

Color Brewer proposes 3 different color schemes. If we intend to encode the following data variables using color, indicate, for each one, what will be the most appropriate color scheme

Client Name      Not Applicable ▴ ▾

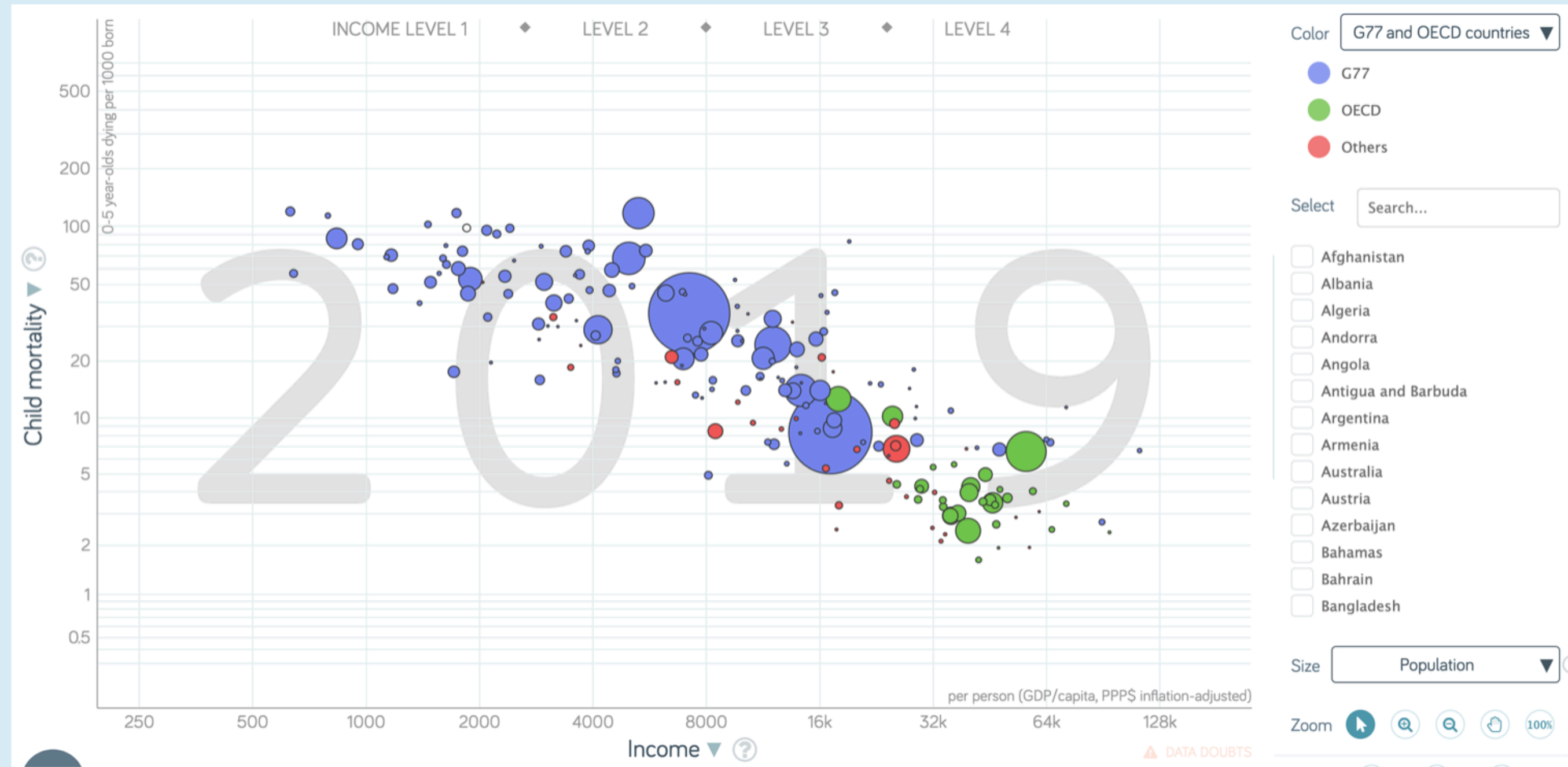
Customer Segment      Qualitative ▴ ▾

Profit      Divergent ▴ ▾

Product Category      Qualitative ▴ ▾

Sales      Sequential ▴ ▾

The graphic bellow, is a typical representative of a **Gapminder** data visualization. On the table below indicate each data variable present in the graphic and the respective visual variable used.



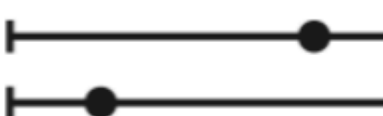
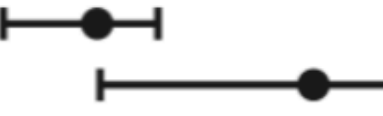



Income	Position-X (Linear)
Population	Size
G77 and OCDE countries	Color
Time	Time
Child Mortality	Position-Y (Log)



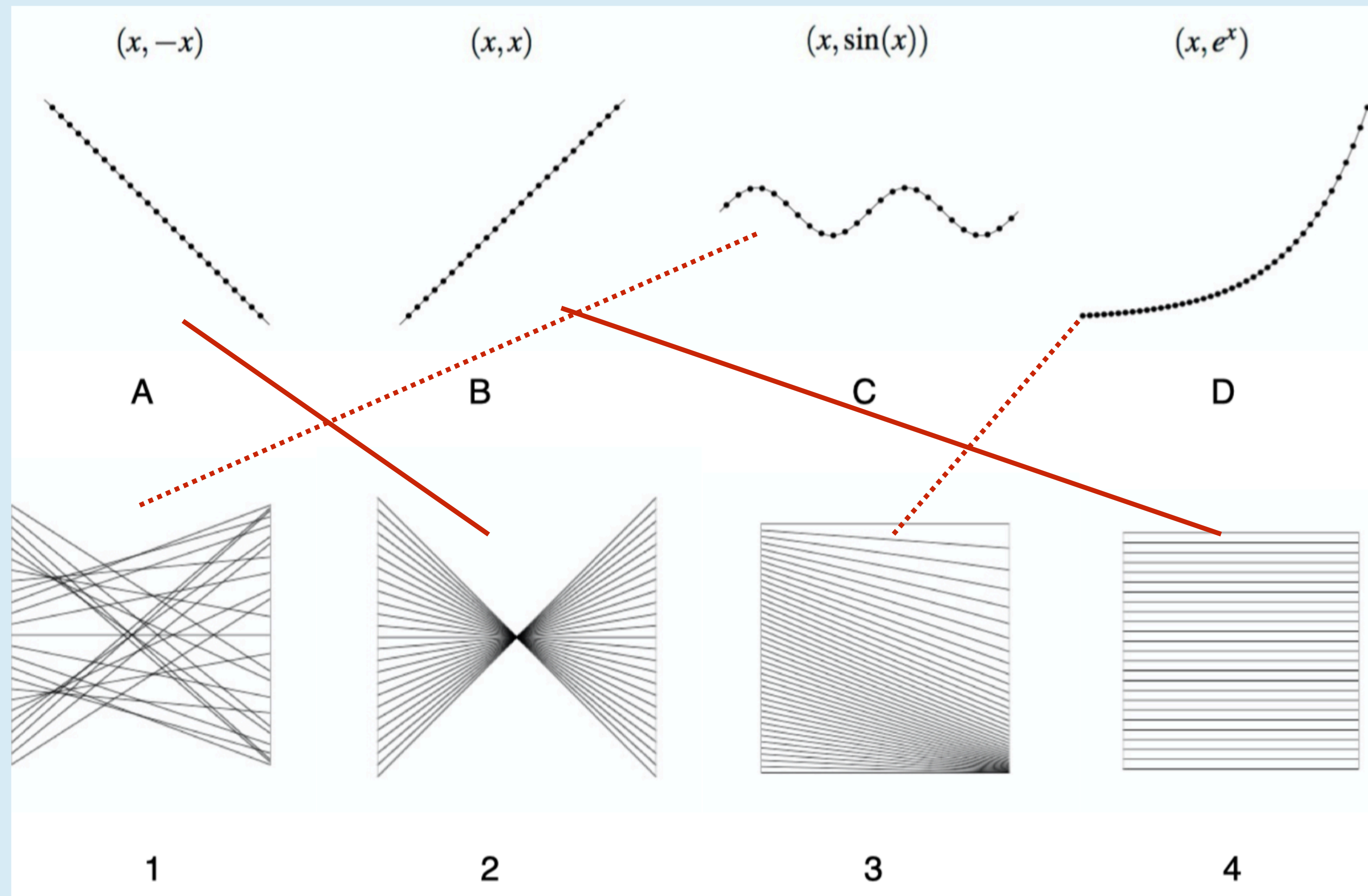
# Order by decreasing effectiveness

Consider the following visual channels to communicate a numeric variable. Order by decreasing effectiveness (best should be first).

- A** Area (2D size) 
- B** Length (1D size) 
- C** Position on common scale 
- D** Position on unaligned scale 
- E** Volume (3D size) 

- A
- B
- C
- D
- E

The figures in the first row correspond to points from a function plotted in orthogonal coordinates. The figures in the second row correspond to some functions (not necessarily on the same order) represented using parallel coordinates. Indicate for each figure from the first row which figure, from the second row, will correspond to it.



B	<input type="text" value="4"/>
C	<input type="text" value="1"/>
D	<input type="text" value="3"/>
A	<input type="text" value="2"/>