

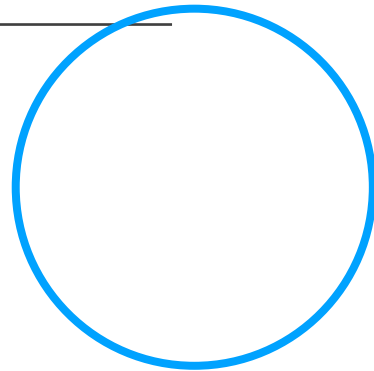


Statistical Graphics and Communication

INTRODUCTION TO DATA VISUALIZATION

KIM HOCHSTEDLER

NOVEMBER 20, 2019





GOALS

I. The importance of data visualization

- **Describe** why data visualization is important.
- **Identify** the main goal of data visualization.

II. General principles of statistical graphics

- **Explain** 3 principles to keep in mind when making graphs.



The importance of data visualization



WHAT DO WE GET FROM TABULAR DATA?



VARIABLE	MEAN	STANDARD DEVIATION	CORRELATION
Salary (thousands of \$)	54.2	16.7	-.06
Driving time to work (minutes)	47.8	26.9	

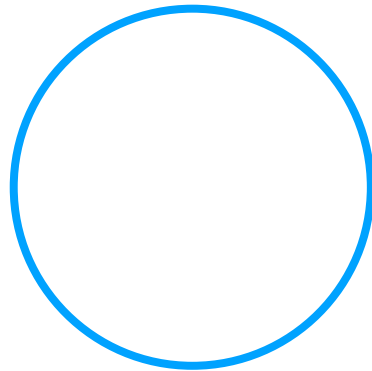
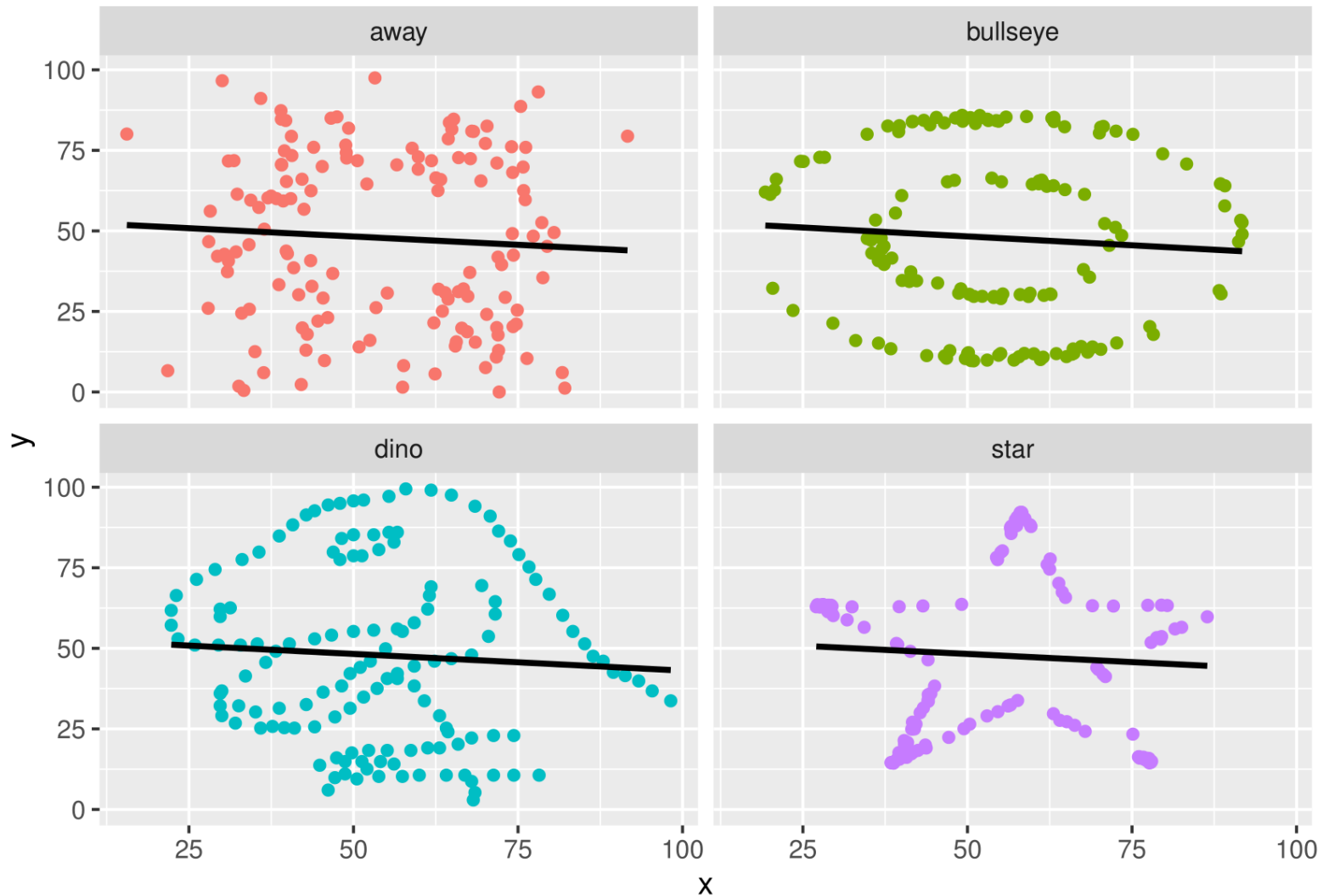


REGRESSION STATISTICS

Intercept = 53.4, Slope = -0.10



WHAT DO THE DATA ACTUALLY LOOK LIKE?



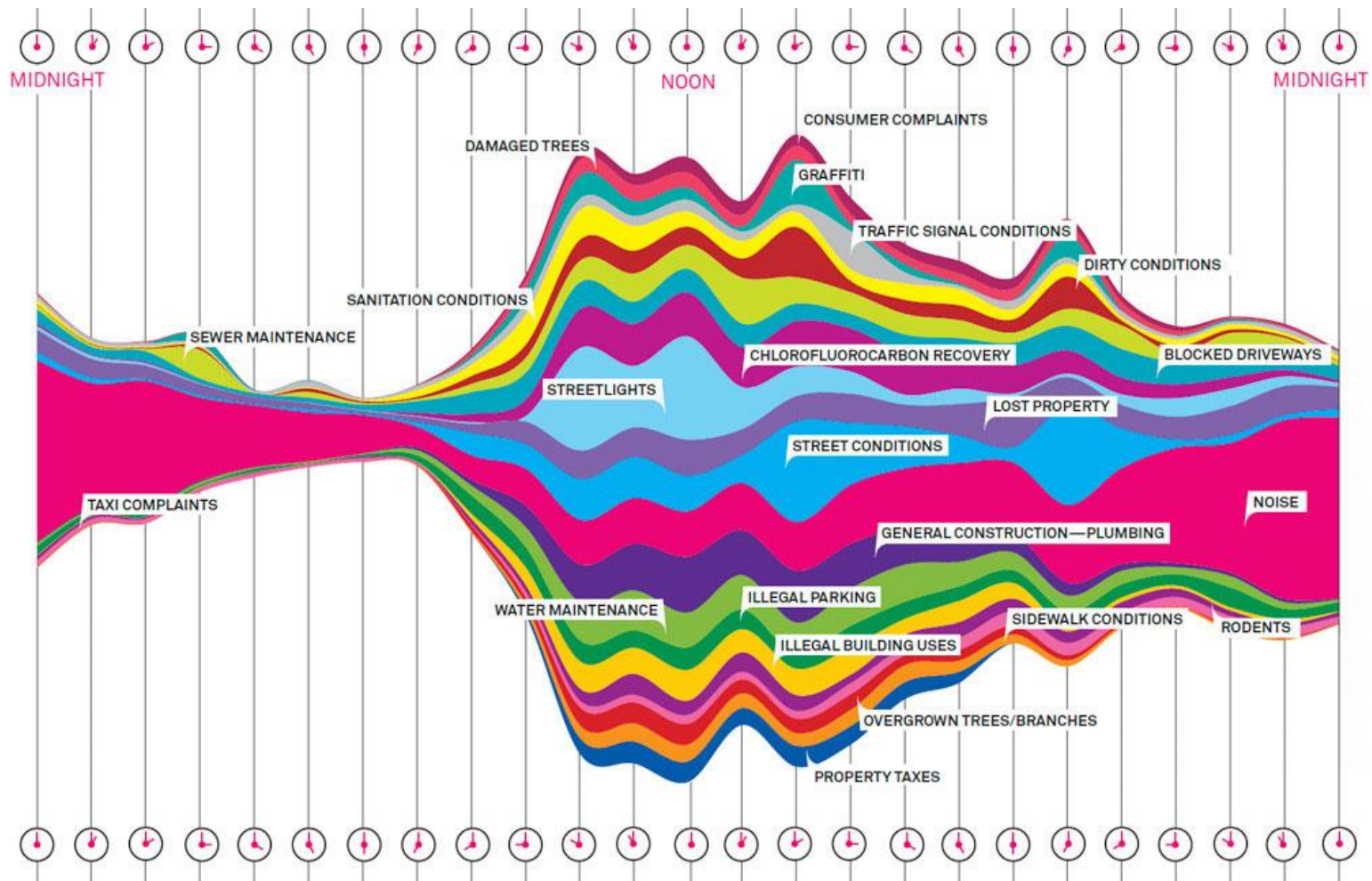
STATISTICAL GRAPHICS

GRAPHICS: visually display data using...



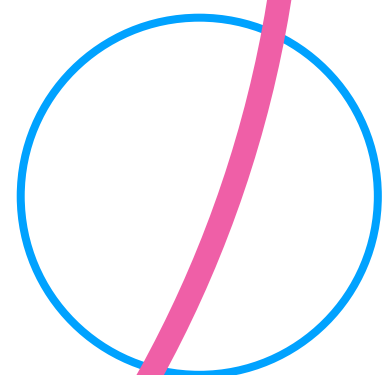
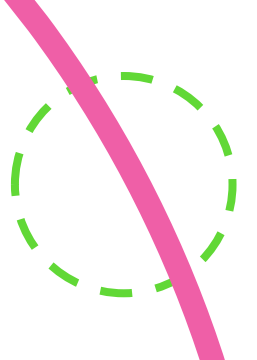
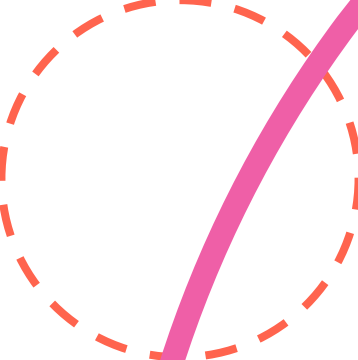
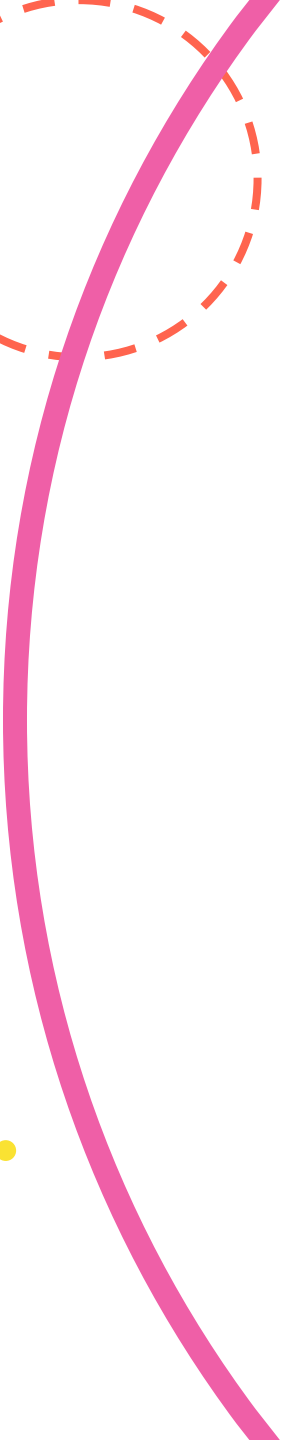
- points
- lines
- coordinate systems
- numbers
- symbols
- words
- color, etc.

-**Balance** aesthetics and information.

-**Goal:** show the data so viewers can make appropriate conclusions.



Ex. Type of 311 calls by time of day in NYC.

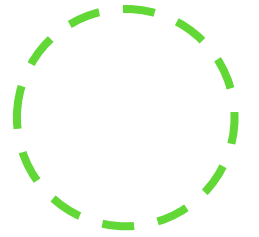


General principles for making graphs

1. GRAPHS SHOULD **STAND ALONE**.
2. AVOID **DISTORTION**.
3. USE **DATA INK** EFFECTIVELY.

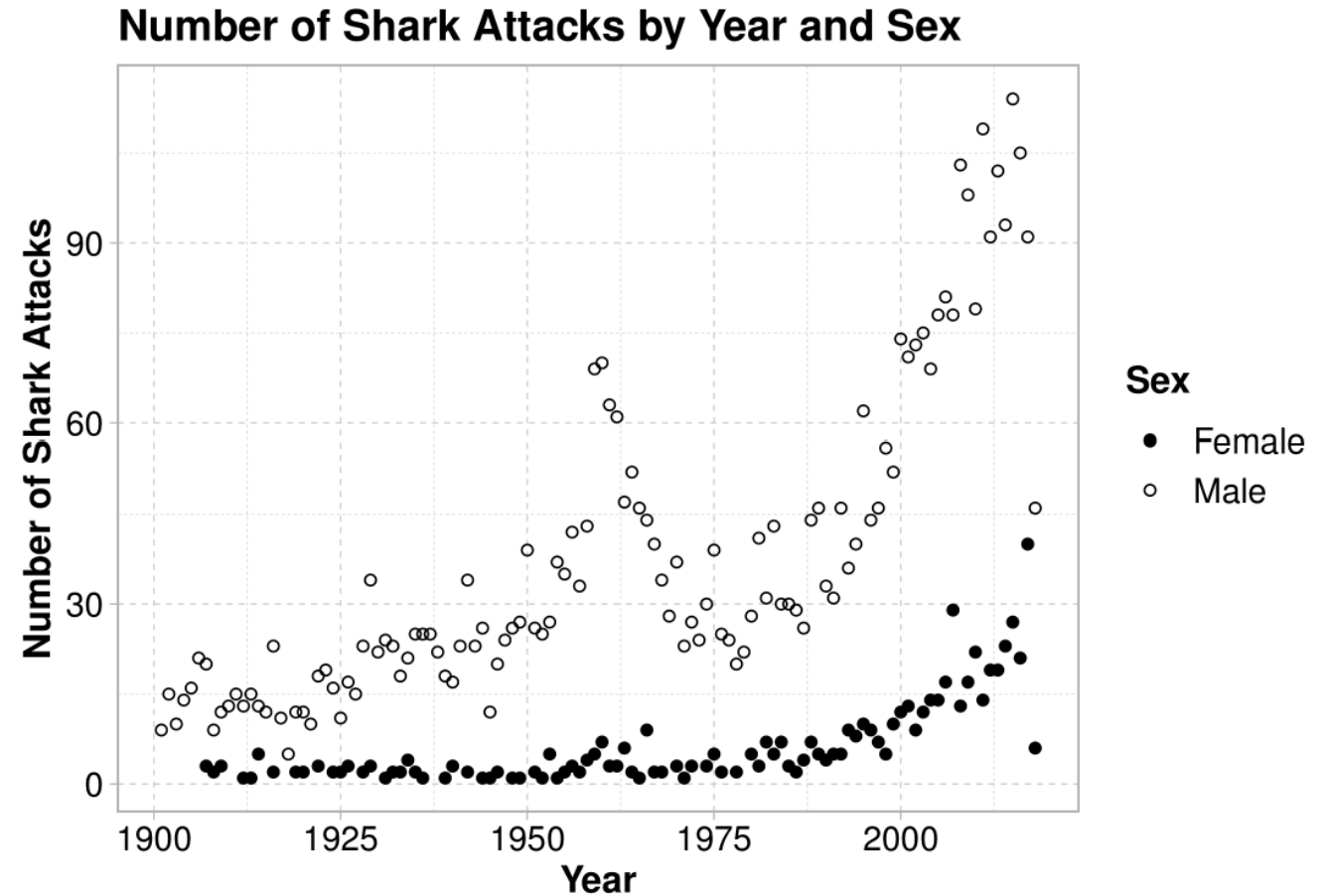
1. GRAPHS SHOULD STAND ALONE.

EXAMPLE



What trends do you notice in this graph?

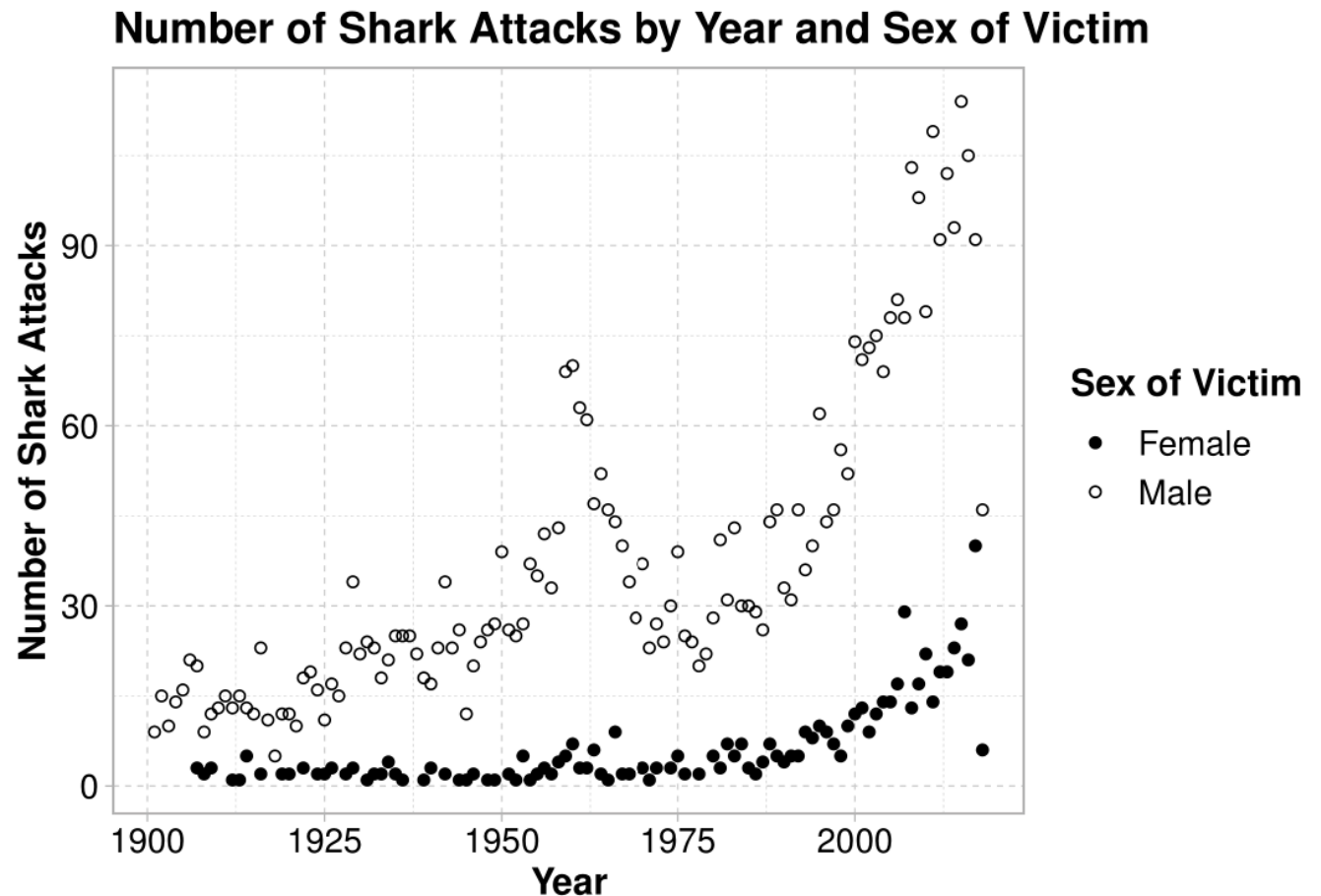
What explanations could you suggest for your observations?



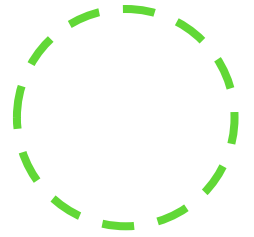
1. GRAPHS SHOULD STAND ALONE.

Readers should be able to draw meaning from the graph in isolation.

- Readers do not need extensive supplementary text to understand the graph.
- Think like an “outsider” when making a graph.



2. AVOID DISTORTION.



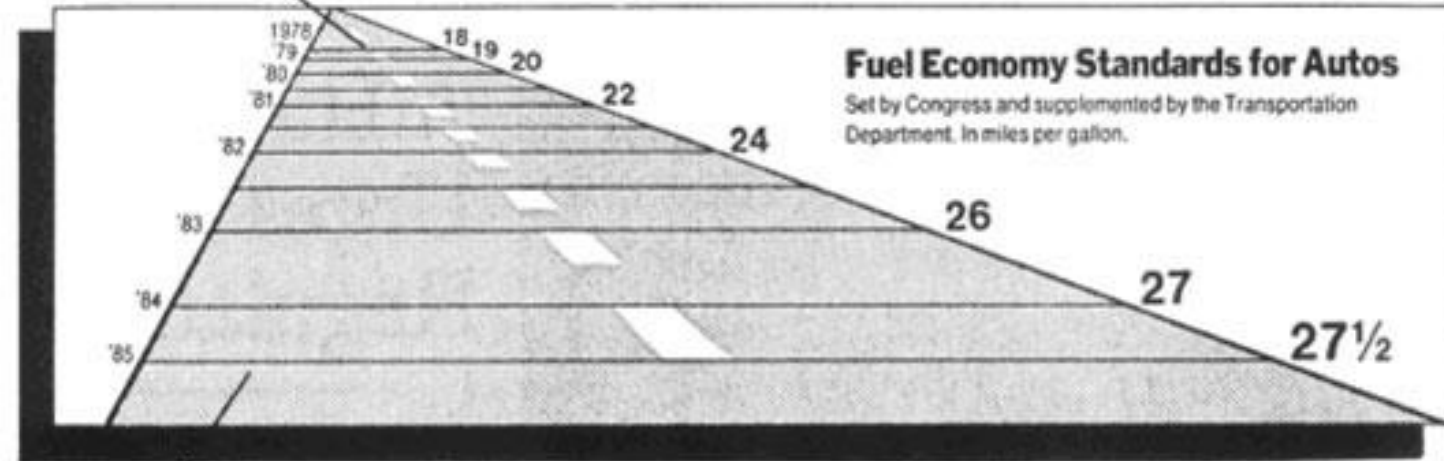
Distortion: graph does not match the data.

We can numerically represent distortion...

$$\text{Lie factor} = \frac{\text{Size of effect in graph}}{\text{Size of effect in data}}$$

- Should be near 1.

This line, representing 18 miles per gallon in 1978 is 0.6 inches long.



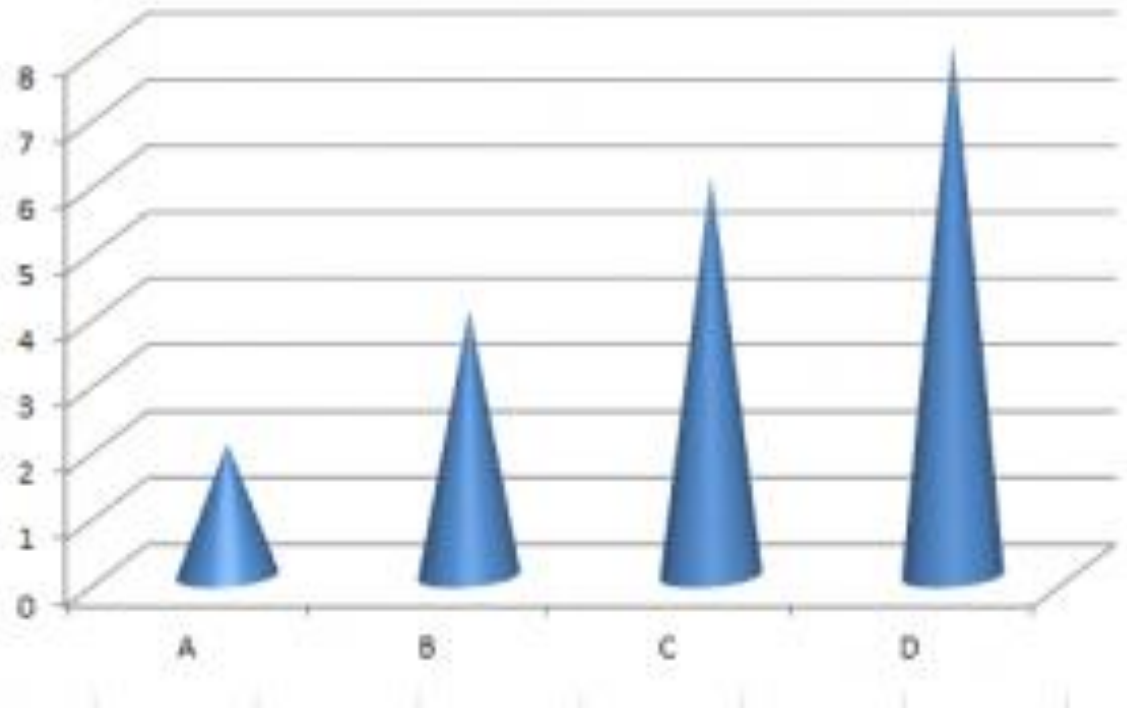
This line, representing 27.5 miles per gallon in 1985 is 5.3 inches long.

$$\text{Ex. LF} = \frac{\% \text{ increase in length}}{\% \text{ increase in data}} = \frac{(5.3 - .6)/.6}{(27.5 - 18)/18} = \frac{783\%}{52.8\%} = 14.83 > 1$$

3. USE DATA INK EFFECTIVELY.

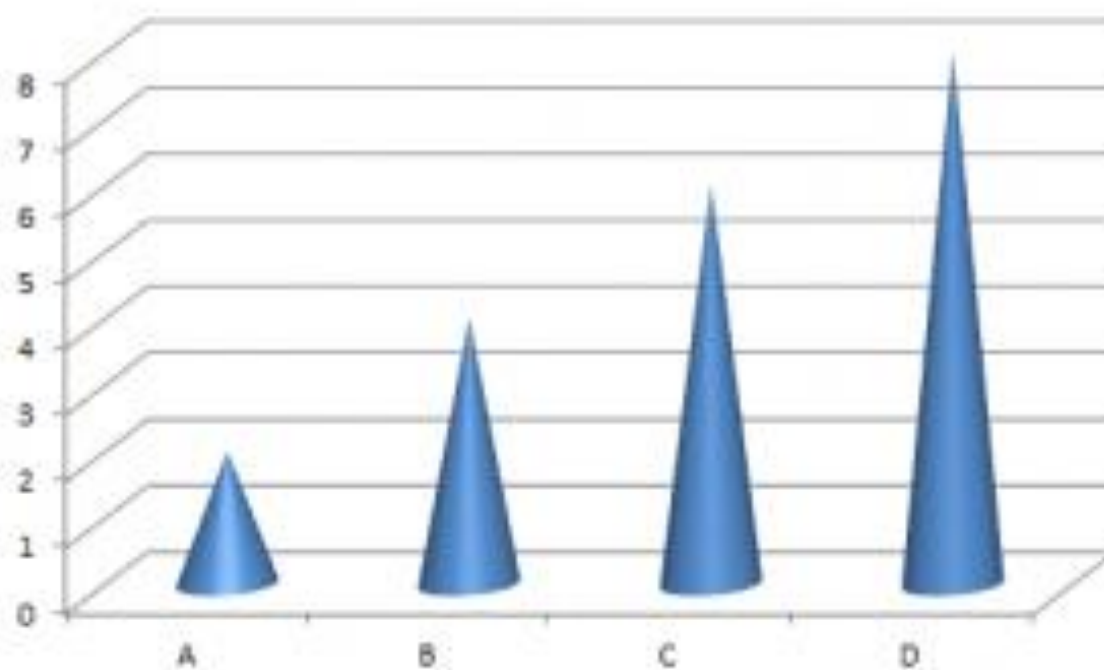
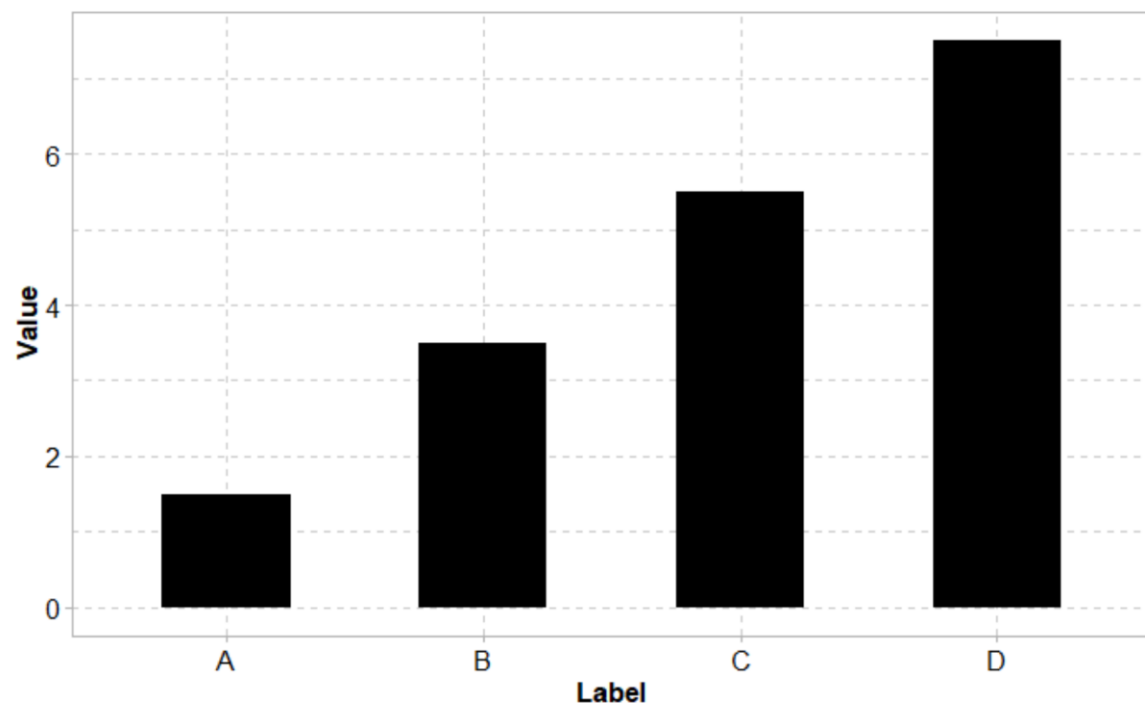
Data ink: “non-erasable” and “non-redundant” core of the graphic.

- Data ink should present information about the data.
- Avoid unnecessary decoration.





3. USE DATA INK EFFECTIVELY.






CONCLUSIONS

It is important to **visualize data** ...
summary statistics are not enough!

Good graphics should...

1. Stand alone.
2. Avoid distortion.
3. Use “data ink” effectively.

Consider these principles in **your own work** and when you see data visualizations in the **media**.



Any questions
or final
thoughts?