Abuses of Visualization

John Keyser

Abusing Visualizations

- Visualizations are known for conveying lots of information.
- Small variations in visualization can convey incorrect information
  - This makes it easier to mislead the viewer
- These are not always intentional, but often are!

Bad Visualizations vs. Abuse of Visualizations

- Will distinguish two types here
- Bad visualizations:
  - Ineffective due to bad choices made
  - Could mislead due to ineffectiveness
  - Color, type of plot, too much/little data, etc.
- Abuse of Visualizations:
  - Choices made that will lead to incorrect interpretation of the data
  - If someone wants to mislead, this is what they do…
- Focus today is on Abuse of Visualizations

Some History

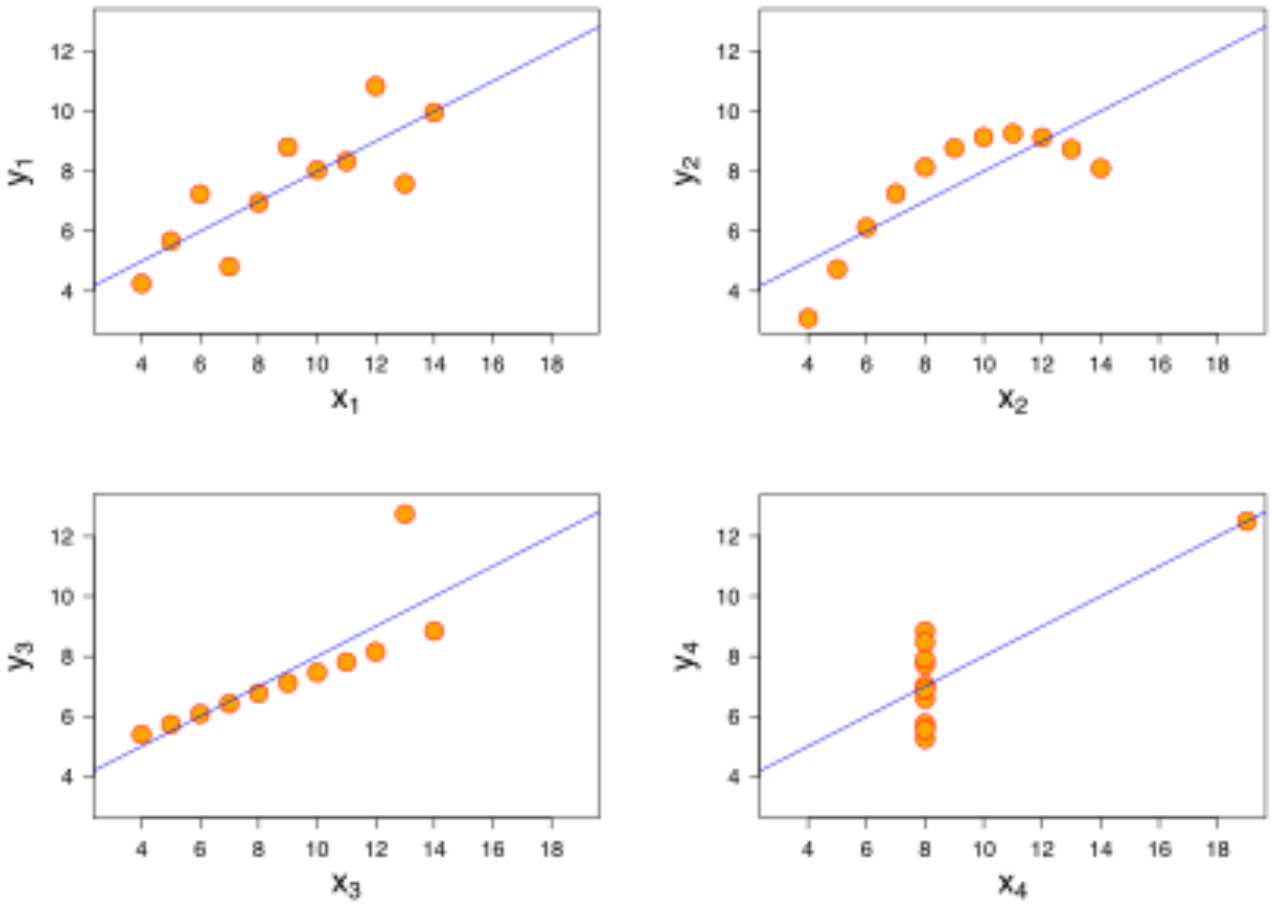
- Famous books:
  - Darrell Huff. *How to Lie with Statistics*. W.W. Norton and Company, New York,1954.
  - Mark Monmonier. How to Lie with Maps. University of Chicago Press, Chicago and London, 1996.

Data Manipulation

- Sometimes the manipulation of the data itself causes issues.
  - Merging/combining/averaging data that should not be.
  - Choosing a histogram bin size too large/small
  - Reporting statistical data that has no meaning or hides the actual data distribution
  - Fitting a curve inappropriately
  - Showing change in derivative instead of value
  - Etc.
- Key: be sure that any manipulation does not destroy the interesting information

Anscombe's Quartet

- Four distributions with the same summary statistics:
  - Mean, variance (x and y), correlation, linear regression
  - F. J. Anscombe, "Graphs in Statistical Analysis," *American Statistician*, vol. 27, Feb. 1973, pp. 17-21.
- See also the "Datasaurus Dozen"



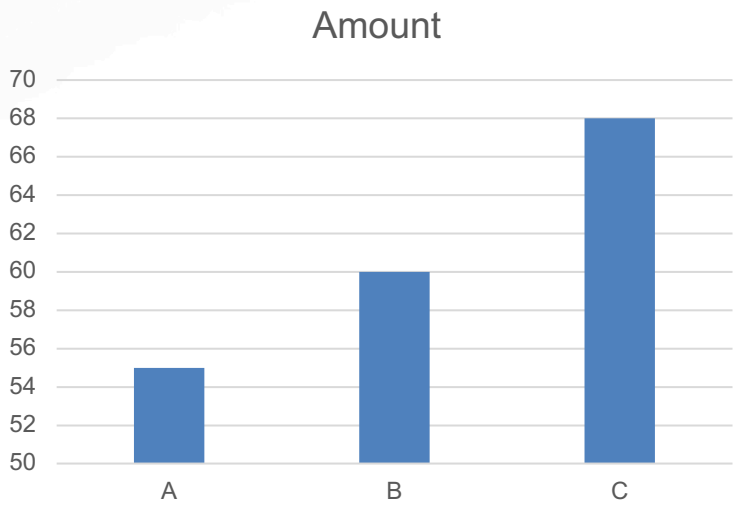
https://en.wikipedia.org/wiki/Anscombe%27s_quartet

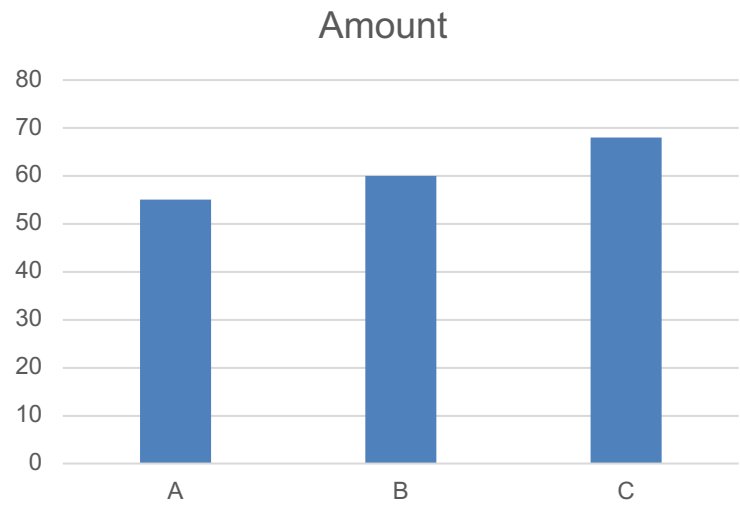
Adjusted Axis Range

- One of the easiest things to do!
- Also one of the most deceptive!!
- And, quite common to find in practice!!!

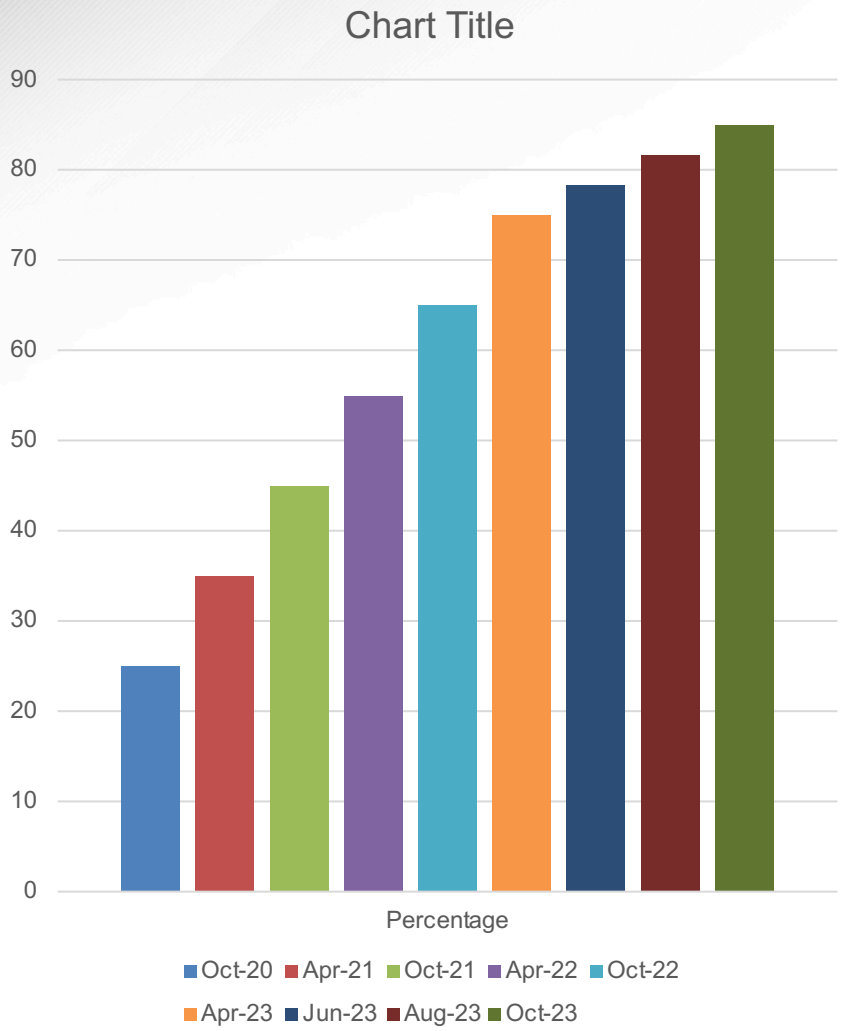
- Starting y-axis at non-zero value
- Logarithmic scale vs. Linear scale
- Use x-axis for ordinal but not numeric values

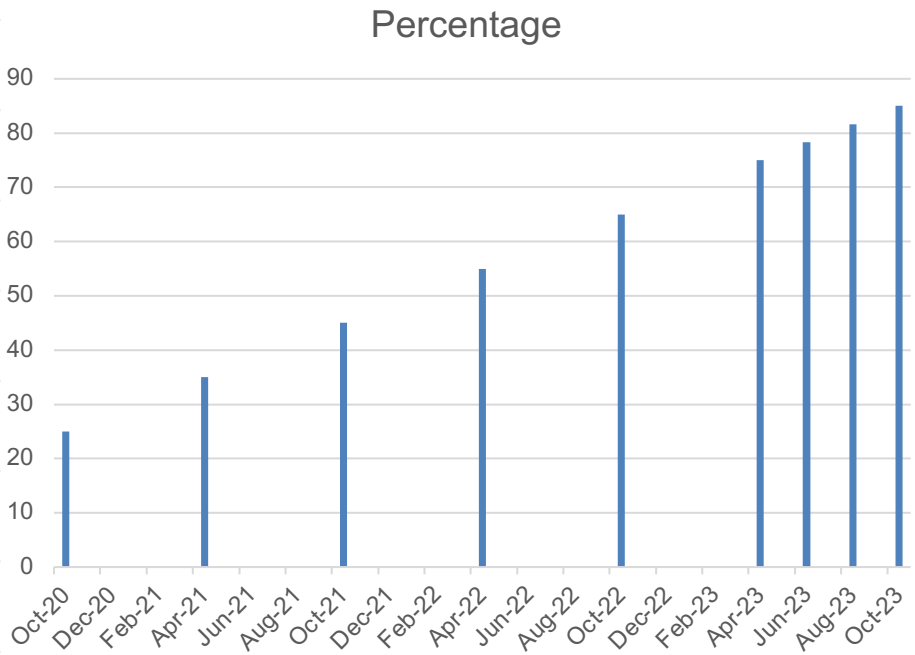
Truncated y Axis





Adjusted x-Axis Range

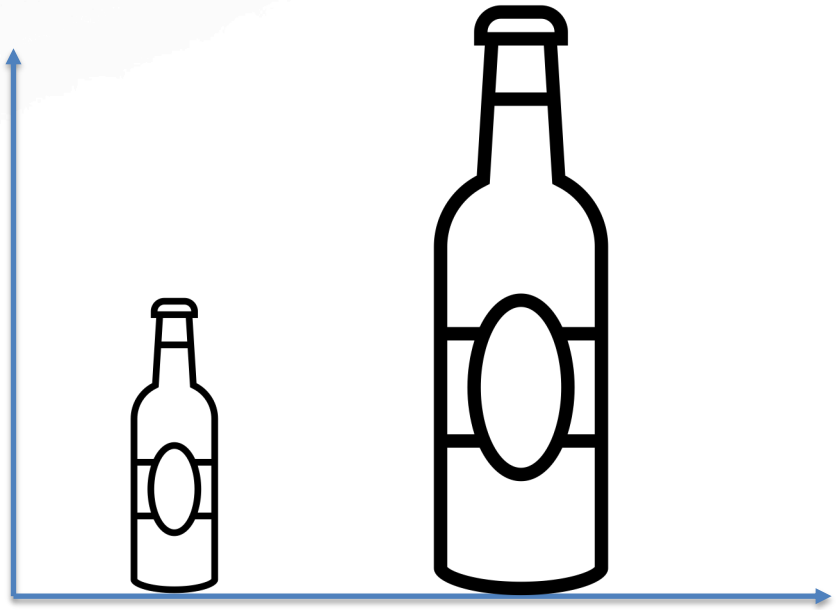




Linear Scale, Area/Volume Appearance

- We perceive amount based on size
  - Area or volume
- Sometimes charts scale a 2D/3D object based on one axis, but object scales

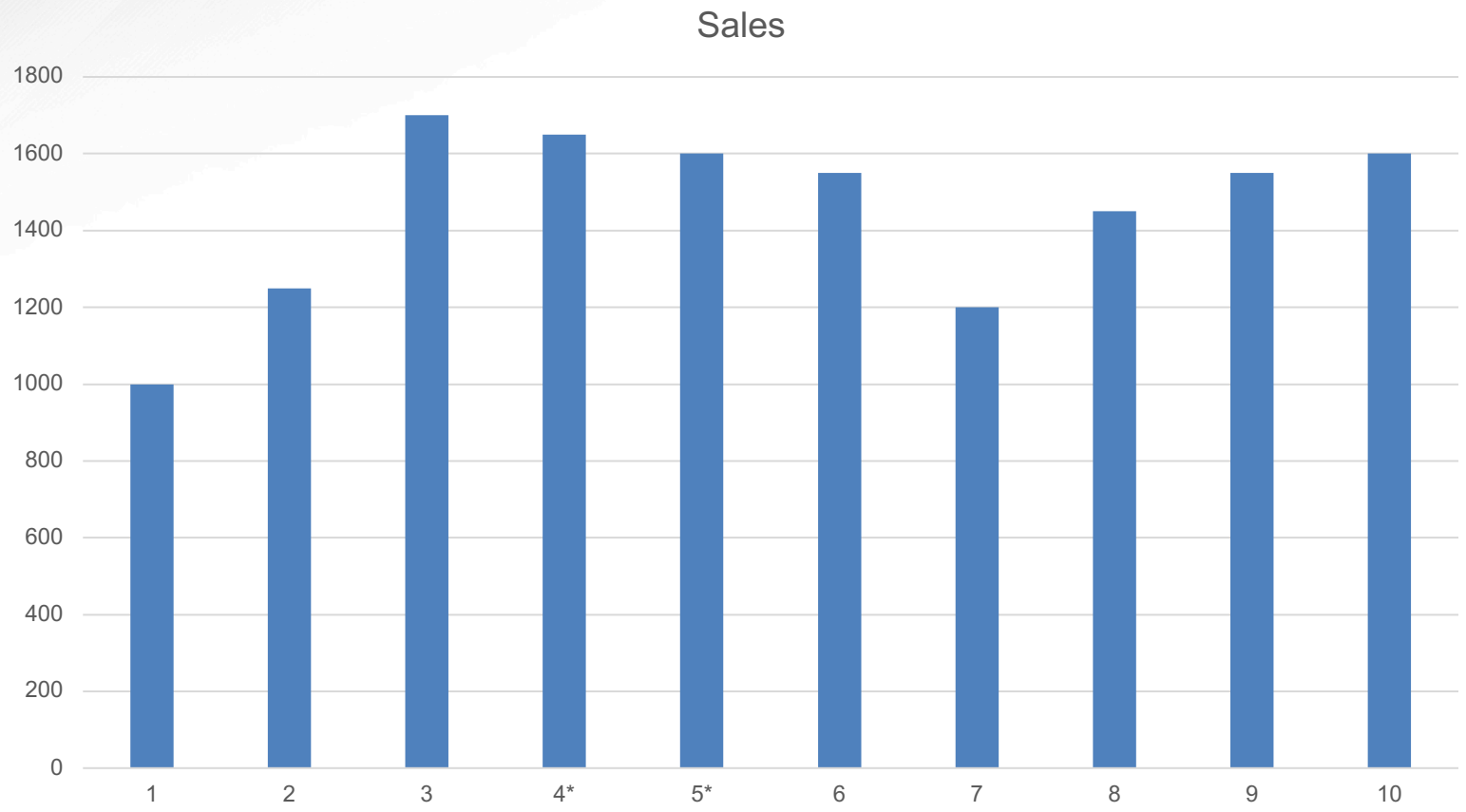
Linear Scale, Area/Volume Appearance



Interpolated/Projected Data

- Data points might not actually exist, but get added in to make the chart/data more “complete”
- Can give the impression of real data where there is not any data, even if labeled appropriately

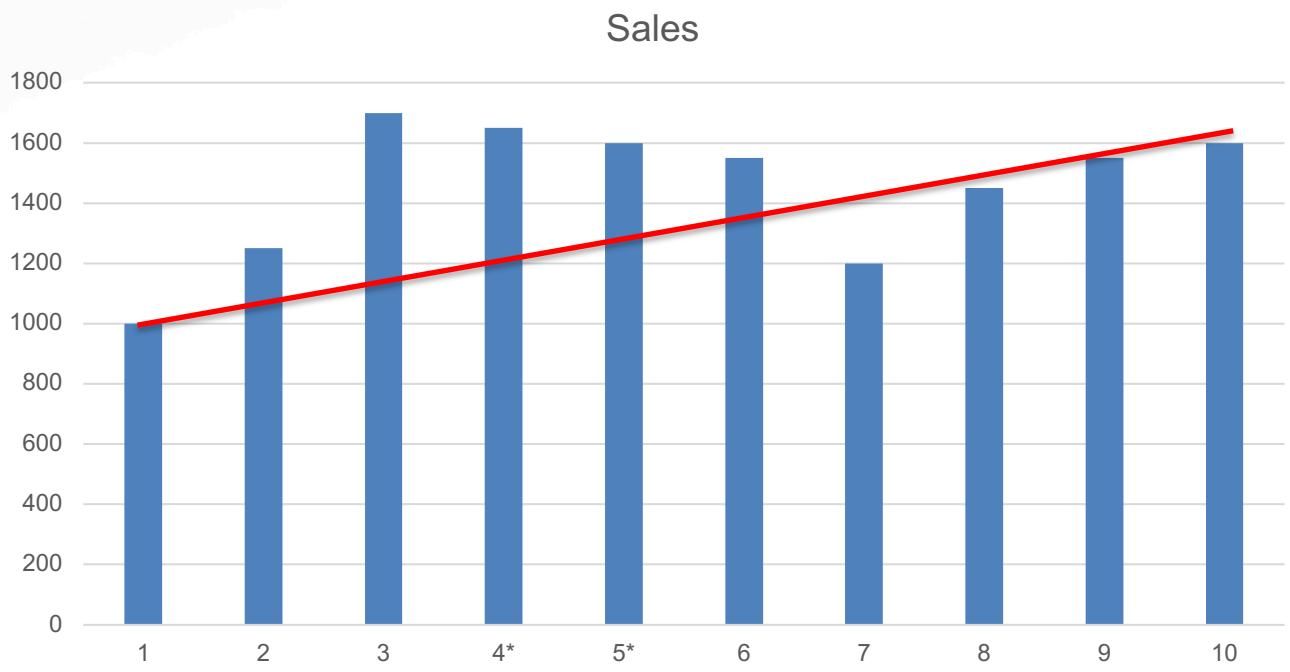
Interpolated/Projected Data



Fitting Curve with Little Basis

- When a curve is shown over the data, it often implies that this is the “base” or accurate value
  - And the data points are just variations from that “true” data
- This curve/line may not have much basis or validity

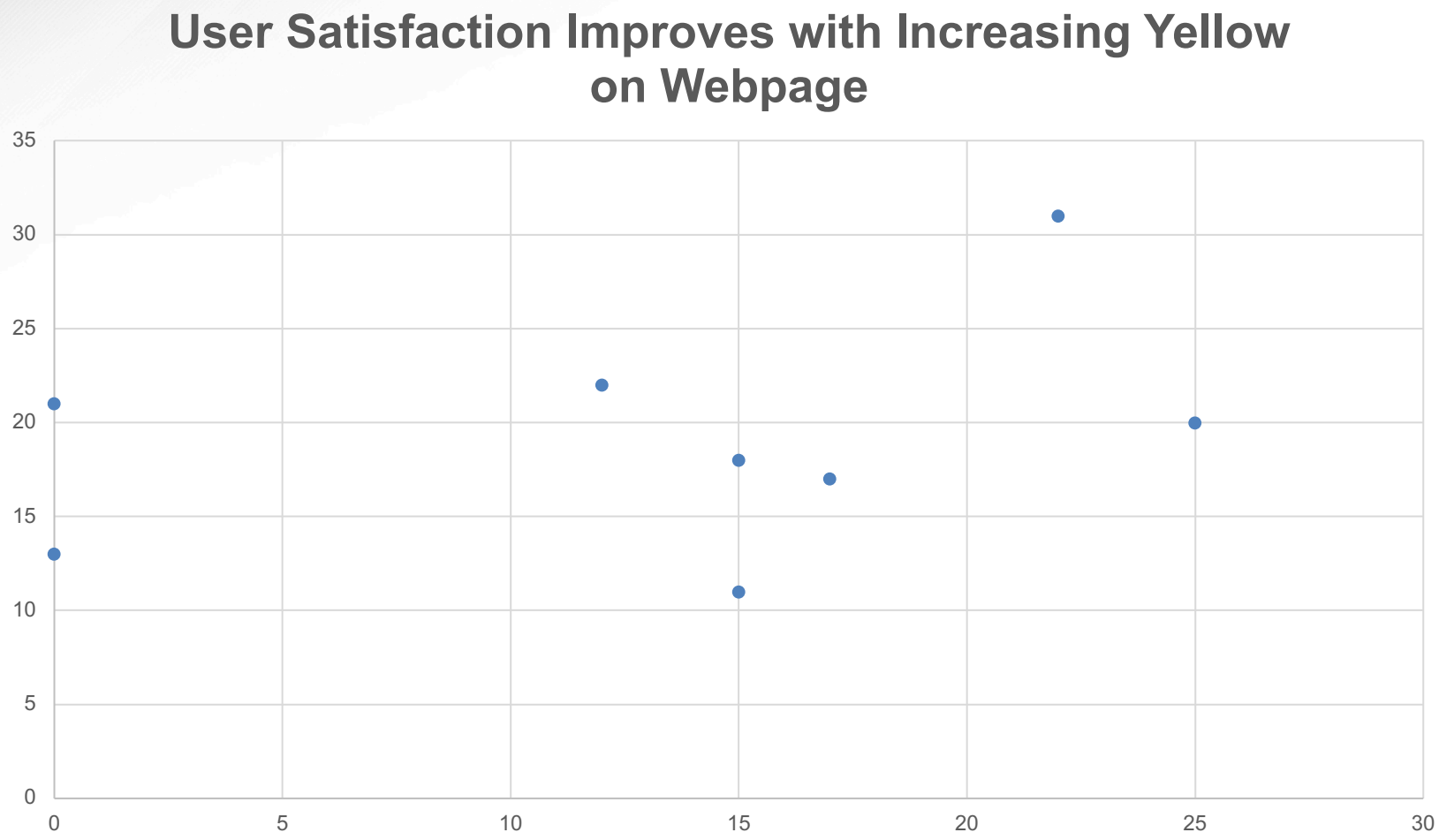
Fitting Curve with Little Basis



Misleading Title/Text

- People can believe a chart shows something that might not really be there, if the title or other text says so.
- One example: stating or implying causation where it does not exist

Misleading Title/Text



Implied Area as a Factor

- In a chloropleth (map with regions colored in by color), the area of a region might not be appropriate
  - e.g. if population-related, the denser cities have less visual impact than less dense rural areas

Implied Area as a Factor