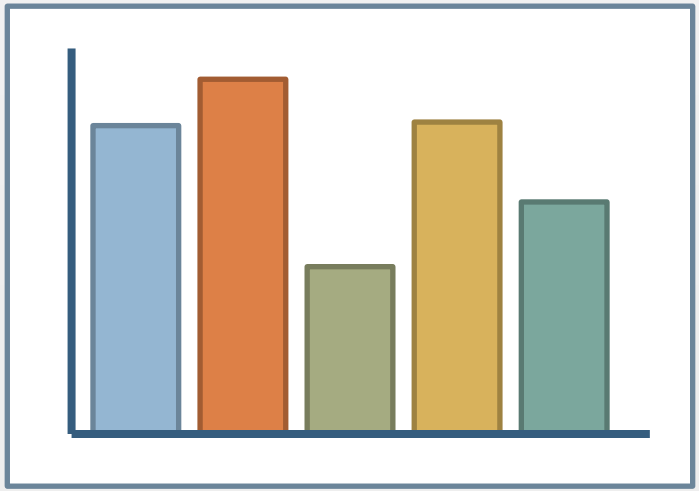
Stacked Graphs

John C. Hart
Department of Computer Science
University of Illinois at Urbana-Champaign

Bar Chart

Quantitative dependent variable

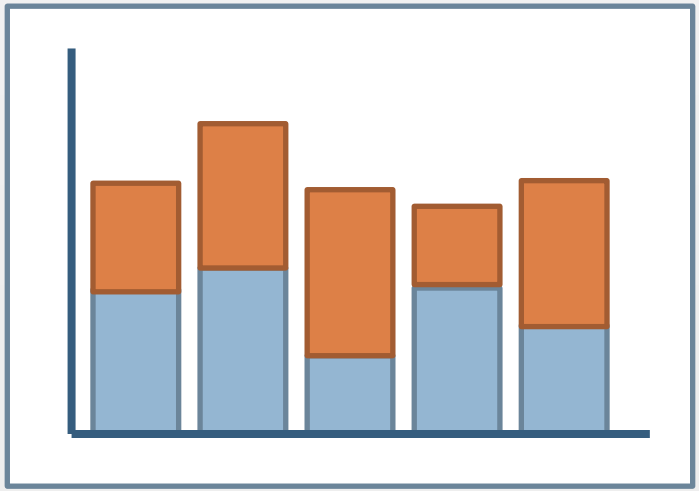


Independent variable

Benefits from both position (top of bar) and length (size of bar)

Stacked Bar Chart

Two (accumulating) q. dep. variables

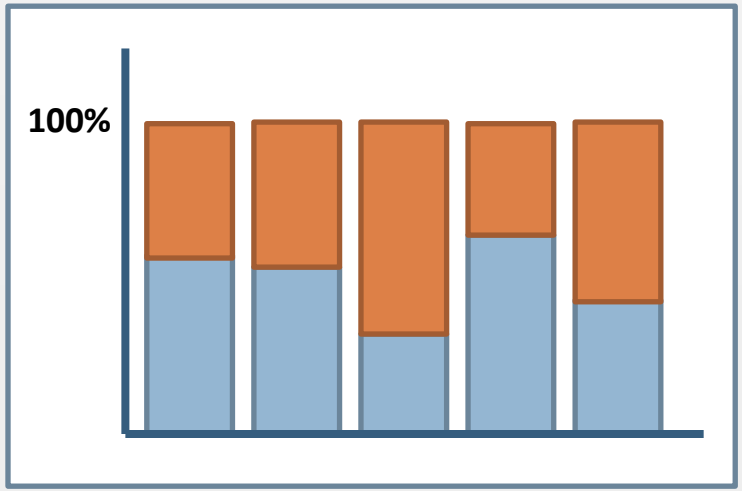


Independent variable

Central limit theorem → as more bars are added, sums will vary less

Relative Stacked Bar Chart

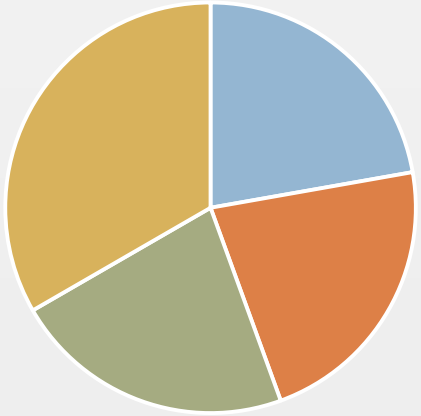
Two quantitative dependent variables



Independent variable

Pie Chart

- Used to indicate relative portions of a quantitative dependent variable of a single dimension
- Maps percentage of total to angle of wedge arc
- Perspective (both distortion and foreshortening) confounds perception of angle



Position

Length

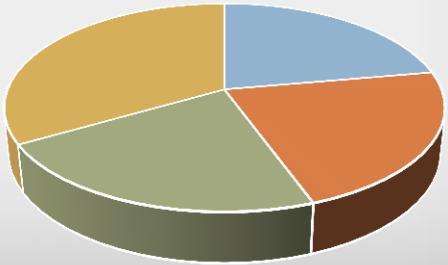
Angle

Area

Volume

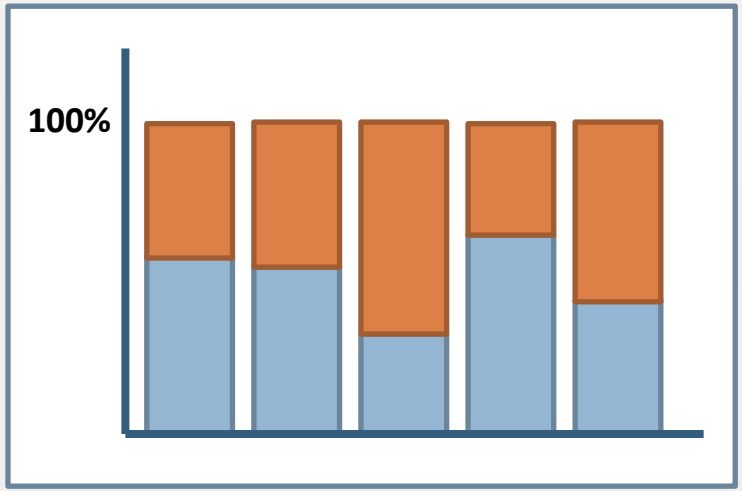
Color

Cleveland & McGill, 1984



Relative Stacked Bar Chart

Two quantitative dependent variables



Independent variable

Position
Length
Angle
Area
Volume
Color

Cleveland & McGill, 1984

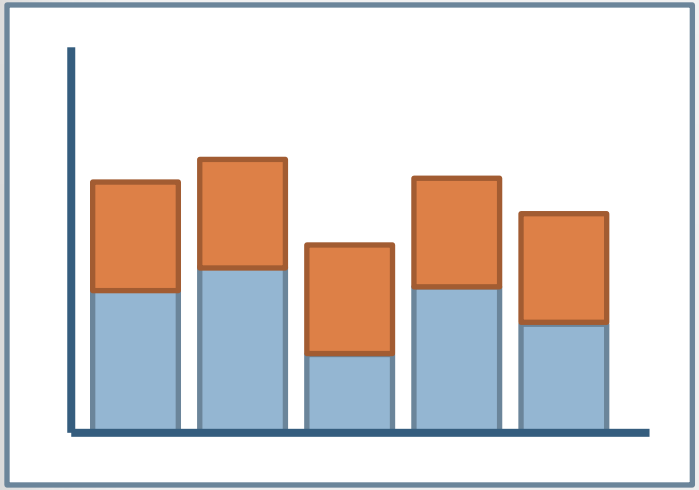
Stacking Order Matters

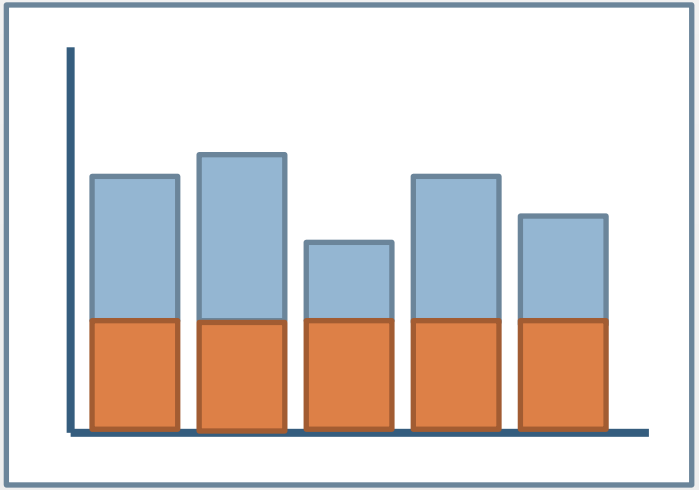
Position

>

Length

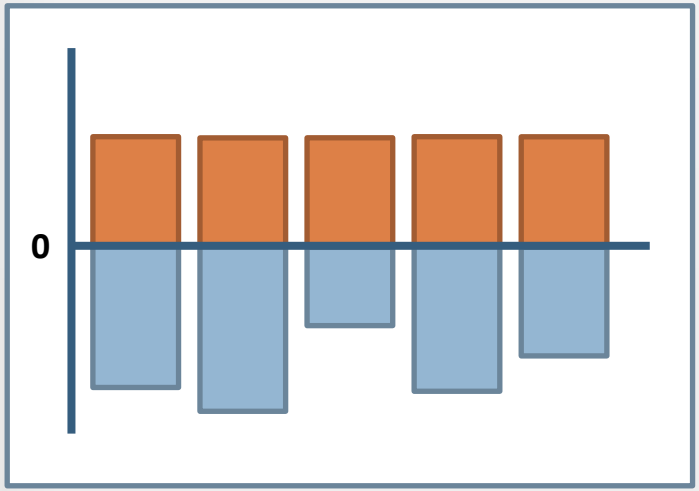
Cleveland & McGill, 1984



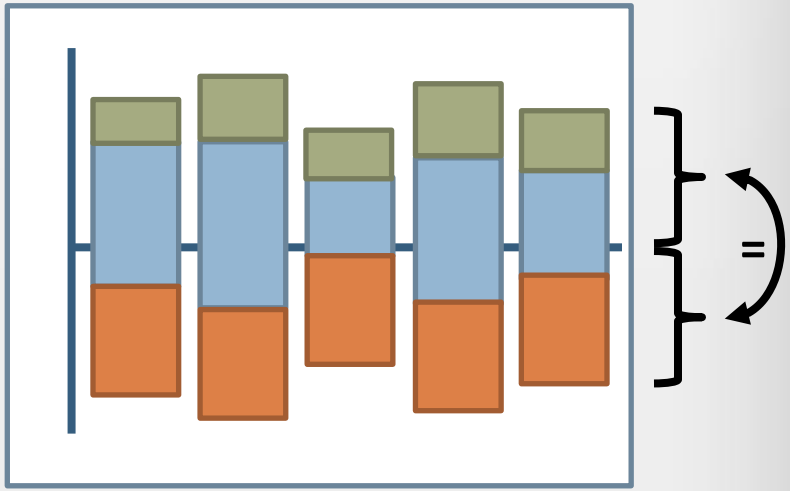


Variance of lower stack elements influences perception of upper stack elements

Diverging Stacked Bar Charts

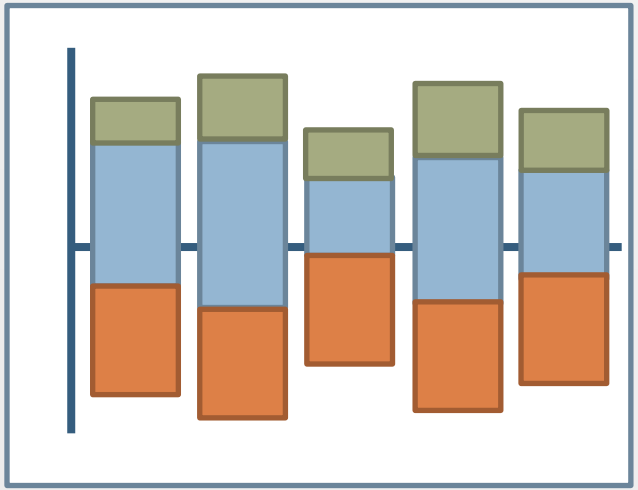


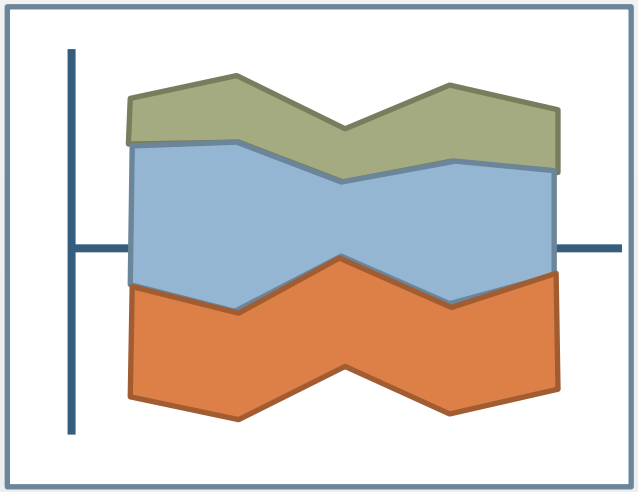
- Benefits from pos. & length
- Only works for two variables
- Negative connotation for lower bars



- Only indicates length
- Works for many variables
- Bar trends can still be obscured by neighboring bar variance

Stacked Bar Charts v. Stacked Line Graphs





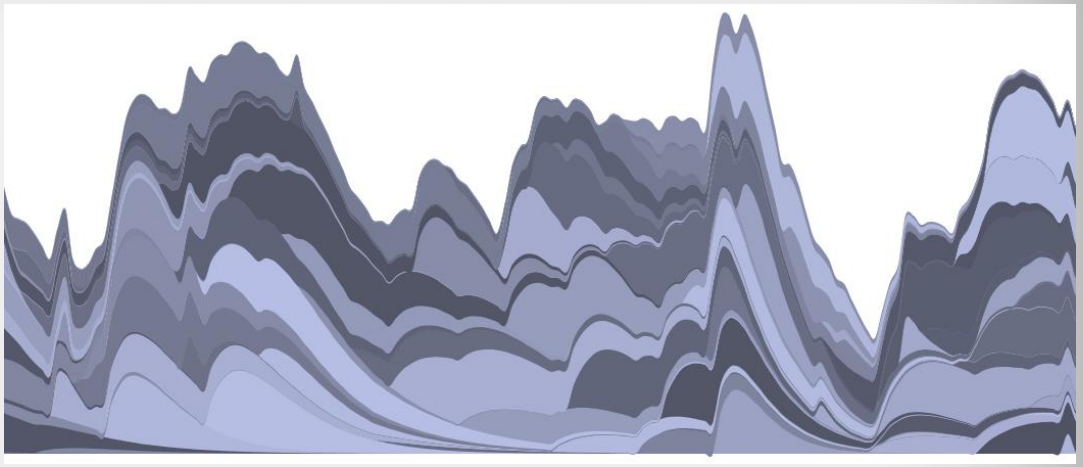
- Appropriate for continuous data over a continuous independent variable
- Can smooth regions using curves instead of line segments

Stacked Graph Layout

- Let g_i be the position of the top of the i'th stacked bar

$$g_i = g_0 + f_1 + f_2 + \ldots + f_i$$

- Setting $g_0 = 0$ results in an ordinary bar chart that distorts data when stacked on varying data underneath



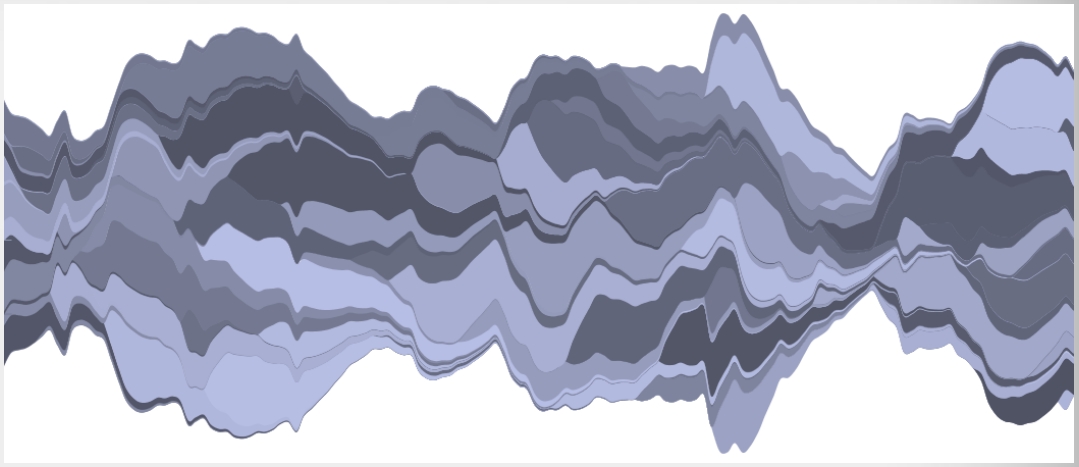
ThemeRiver Layout

- Let g_i be the position of the top of the i'th stacked bar

  $g_i = g_0 + f_1 + f_2 + \ldots + f_i$

- ThemeRiver centers the bar chart on the horizontal axis by setting

  $g_0 = -½ (f_1 + f_2 + \ldots + f_n)$



- Minimizes the girth of the chart ($g_0^2 + g_n^2$) and the top and bottom slopes ($g_0'^2 + g_n'^2$)

- Havre, S., Hetzler, B., Nowell, L. ThemeRiver: Visualizing Theme Changes over Time. *Proceedings of the IEEE Symposium on Information Visualization, 2000*

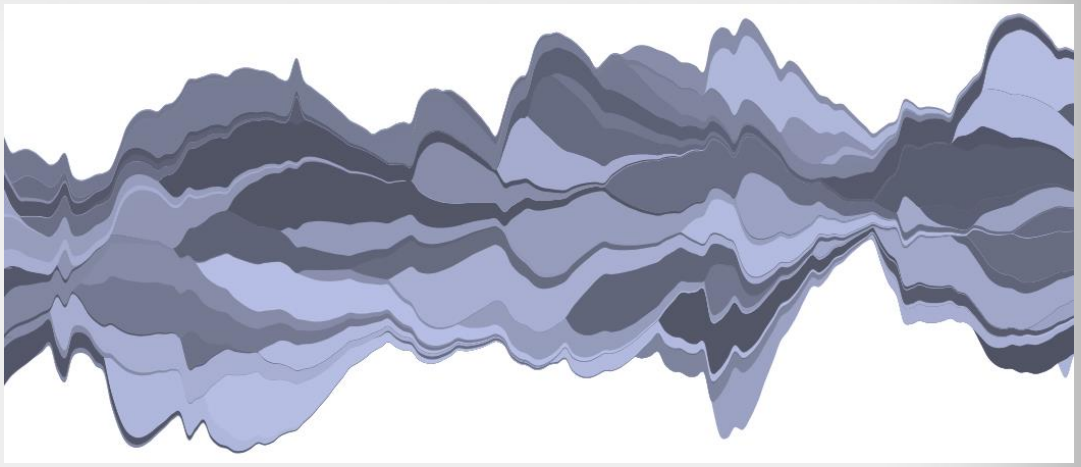
Streamgraph Layout

- Let g_i be the position of the top of the i'th stacked bar

$$g_i = g_0 + f_1 + f_2 + \ldots + f_i$$

- Streamgraph sets the base at

$$\boldsymbol{g_0 = -\frac{1}{n+1}\sum_{i=1}^{n}(n-i+1)f_i}$$



(actually uses a weighted version, but harder to evaluate)

- Minimizes the "deviation" ($\Sigma\, g_i^2$) and the "wiggle" ($\Sigma\, g_i'^2$)
- Byron, Lee, and Martin Wattenberg. "Stacked Graphs – Geometry & Aesthetics." IEEE Trans. On *Visualization and Computer Graphics* 14(6), 2008, pp. 1245-1252.

Streamgraph Ordering

- Compute total weight w_i of each series i (sum of values of each datapoint)
- If $(w_1 + \ldots + w_{n/2}) > (w_{n/2+1} + \ldots + w_n)$, then add next series to bottom, otherwise add next series to the top
- By adding new series at bottom (f_1) or top (f_n), new data is introduced near high-contrast silhouette where it is better noticed, and fades toward middle