Data types in visual analytics

* Numeric
  + Measure quantity as number
  + Can be continuous/discrete
* Categorical
  + Describe a quality/characteristic
  + Can be Ordinal
    - Ordered, ranked, size
  + Can be nominal
    - Not organized into a logical sequence
    - Gender, business type, eye color, brand
  + Usually plotted as bar charts or pie charts
* Text
* Time series
* Graph and networks
* Hierarchies

Data not a number

* Images
* Videos
* Text
* Web logs
* …
* Requires feature analysis to turn these abstract things into numbers
  + Apply the result as usual
  + Keep reference to origin data to maintain native domain

Sensor Data

* Large scale
* Feature analysis
  + FT, FFT, WT, FWT

Image Data

* Array of pixels
* Feature analysis
  + Value histogram
  + Encode into a 256-D vector
  + Gradient histogram
  + Visual words
  + FFT, FWT
  + Scale invariant Feature Transform
  + Bag of Features

Video Data

* A time series of images
* Feature analysis
  + Same as image data

Text Data

* Often raw and unstructured
* Feature Analysis
  + Term-document matrix
  + Word Embedding

Web logs

* Usually text strings in a pre-specified format
* Easy to convert into categorical and numerical attributes

Network Traffic

* Network packets
* Features cant be extracted
  + Number of bytes transferred
  + The network protocol used
  + IP ports used

Hierarchies

* Tree with quantities
* Partition of unity
* Information flow
* Force directed layout

Convert Numeric to Categorical Data

* Method 1
  + Divide numerical data into equal range interval
  + Each range/bucket has same width
  + Problem
    - Data could be unevenly distributed within a bin
* Method 2
  + Divide numeric data into equal depth ranges
  + Same number of samples in each bin
  + Problem
    - Extra storage needed: must store the start/end value for each bin
* Method 3
  + Entropy based binning
  + Divide numeric based on the calculated entropy

Anti-aliasing

* Sample at a higher rate
* Or smooth the signal before sampling it (filtering)
* Result
  + Sharper image
  + But has jaggies

Smoothing

* Slide a window across the signal
  + Stop at each discrete sample point
  + Average the original data points that fall into the window
  + Store this average value at the sample point
  + Move the window to the next sample point
  + Repeat
* Result
  + A bit blurred

Bar vs Histogram

* Histograms
  + Bars show the frequency of numerical data, cannot be reordered
  + Elements are grouped together, considered as ranges
  + Width of bars need not be the same
* Bar
  + Uses bars to compare different categories of data
  + Comparison of discrete variables
  + Can be reordered
  + Width of bars need to be the same
  + When number of bars >=50, change to line chart