590 DV Lecture 4 Note   Feb 11ᵗʰ, 2019

Warm-up Activity:

An analysis and discussion about a visualization of distances

between American cities.

Method: using friendship links between pairs of anonymous

Facebook users.

Can type in specific county to see its relations or connections to

other counties or states.

Today's topic:

Data types in a Visualization and Distributions

The way we express a data point can be a method to classify data

Data that is continuous in one or more dimension can enable

additional operations (e.g. take a slice image)

A given set of points -> a continuous distribution:

Line graph: too general

Histogram: better, more meaningful trendline

    Drawbacks: missing data would cause a misleading dip

    Solvation: Non-uniform binning, combing adjacent bins until

            every has at least one point

    --------- reduces the artificial drop, but also blurs out the original

            trend

Non-uniform bins require time for search

Statistics:

    Summation – describe total quantity

    Arithmetic average – describe mean quantity

    Weighted average – describe mean as a function of other

            variable