



The University of Chicago
Department of Statistics
Seminar Series

LELAND WILKINSON
Department of Computer Science
University of Illinois at Chicago

Scagnostics

MONDAY, May 19, 2008 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

Scagnostics is a Tukey neologism for the term *scatterplot diagnostics*. Scagnostics are characterizations of the 2D distributions of orthogonal pairwise projections of a set of points in multidimensional Euclidean space. These characterizations include such measures as density, shape, outliers, and texture. Tukey never published the idea. Paul Tukey discussed it briefly at an IMA workshop on visualization in 1989.

We introduce a set of scagnostics measures based on graph theory and we analyze their empirical distributions and performance. Our analysis leads to a restrictive set of criteria that must be met in order to have scagnostics measures that can be used effectively in exploratory data analysis.

We will present several visualization applications that employ scagnostics. One is an interactive, exploratory framework that links scagnostics with scatterplots and scatterplot matrices. The other is an automated visualization system for displaying heterogeneous data sources.