**CoxJones 1981 Exploratory Data Analysis**

In Exploratory Data Analysis, attempts are made to identify the major features of a dataset of interest and to generate ideas for further investigation. While in Confirmatory Data Analysis, attention is focused on model specification, parameter estimation, hypothesis testing and firm decisions about data. This distinction made by the statistician J.W. Turkey, is essentially that between descriptive and inferential statistics.

Four themes are apparent on Exploratory Data Analysis;

1. Displays,
2. Residual,
3. Transformations and
4. Resistance.

1. Displays: Plot both your data and the data and the results of data analysis is one of the basic attitudes of EDA. Displays reveal the major features of data, help in the production of ideas for further investigation and are useful in checking assumptions. Graphical inspections allows the identification of outliers, nonlinearities, discontinuities, skewness and other characteristics of the data which may make or mar the analysis.
2. Residuals are the remainders left after a model after(any kind of summary description, from something simple like a measure of level to something more complex like a multiple regression) has been fitted to data. Graphical display is the most useful weapon available for analysis of residuals.
3. Resistance: much theory behind statistical methods is concerned with optimal procedures, which are best according to specific criteria if particular assumptions about generating processes are satisfied. It is desirable that procedures be robust and work well under a variety of conditions not just under idealized specific conditions for which they have been proven optimal. Robustness is especially important in Exploratory work since it would be foolish to be confident that an unexamined set of data satisfies assumptions about behaviour.
4. Transformation of geographical data has usually been motivated by an inferential approach, and particularly by the idea that hypothesis testing requires data drawn from Guassian(Normal) distribution. One common aim of re-expressing variables has been normalization of frequency distribution. The degree of success of any transformation may be assessed by inspection of histograms or probability plots, calculation of skewness and kurtosis or performance of some distribution free test such as chi-square.

Smoothing is an approach used in EDA of special interest to geographers who are commonly faced with data in the form of spatial or time-series. It is appropriate if data series may be regarded as a mixture of smooth and rough components, whereby

**data = smooth + rough**

for example, data may be regarded as 'signal' mixed with ' noise', 'true values' mixed with 'measurement error', ' long-term trend' mixed with 'short-term fluctuation', or as 'regional trend' mixed with 'local deviation'. Popular smoothing methods used in quantitative geography can be classified as either linear function fitting or local linear smoothing. In linear function fitting we adopt a precisely specified model of the general form

**data = linear function + stochastic error**

Inference: statistical methods in geography appears to be based, at least implicitly with views that descriptive statistics are rather obvious, if not trivial, while inferential statistics are both more challenging and more valuable intellectual.