Visualization of Summary Statistics and Uncertainty

Kristin Potter, Joe Kniss, Richard Riesenfeld, and Chris R. Johnson

Scientific Computing and Imaging Institute University of Utah

> EuroVis June 9, 2010



2010 JUNE 9 - 11 Bordeaux, France Eurographics/IEEE VGCC Symposium on Visualization





Data & Uncertainty

- Larger, more complex datasets common
- Error, accuracy, confidence level
- Scientific data are incomplete without indications of *uncertainty*



Uncertainty Visualization

- Visually depict uncertainties
- Faithfully present data
- Improve visualization as a decision making tool
- Often displayed as mean & standard deviation (variance)



Mean & Variance is Not Enough

- Standard deviation only gives a measure of data variation
- Mean may not be a valid data value





Traditional Display

- Boxplots
 - Show range of data
 - Quartile range, including median
 - Outliers





Visual Modifications

- Refinement for aesthetic purposes



 ★ Mary Eleanor Spear. Charting Statistics. McGraw-Hill, 1952 John W. Tukey.
 Exploratory Data Analysis.
 Addison-Wesley, 1977.

Edward Tufte, The Visual Display of Quantitative Information. Graphics Press, 1983.



- Density indications
 - Use the box sides to encode density information



★ Yoav Benjamini.
 Opening the box of a boxplot.
 TAS, 42(4) ,1988.

 W. Esty and J. Banfield. The box-percentile pot. JSS, 8(17), 2003.

J. Hintze, and R. Nelson.
 Violin plots.
 TAS, 52(2), 1998.



- Data Characteristics
 - sample size, confidence levels



Robert McGill, John W. Tukey, and W.A. Larsen, Variations of box plots. The American Statistician, 32(1), 1978.



Additional Statistics skew, modality



C. Choonpradub, and D. McNeil. Can the box plot be improved? Songklanakarin Journal of Science and Technology, 27(3), 2005,



The Summary Plot

- Augment boxplot with numerous display techniques
- Emphasize characteristics other than mean/variance
- Indicate quantity & location of uncertainty





Anatomy





Abbreviated Box Plot

- Visual reduction of the box plot
- Minimize density assumptions
- Outliers not removed





Moment Plot

× +++

- Statistical measures of feature characteristics
- Signature similar to boxplot
- Can express features hidden by boxplot (e.g. asymmetry)







Density Plot

- Redundantly encode density through colormap and width
- Symmetric display on either side of plot
- Type of estimator influences display

 Kernel Density Estimation Emmanuel Parzen.
 On estimation of a probability density function and mode. 1962.





? X

Distribution Fitting

Distribution Fitting

- Fit to canonical distributions from a library
- Find a best fit
- Or fit to a chosen distribution



Summary Dialog

Data



2D Box Plots

- Data with correlation relationships
- 2D+ statistics not trivial





Variations on 2D boxplots - 1

RangeFinder Plot

- 1D boxplot in each dimension
- independent in each dimension



S. Becketti, and W. Gould, Rangefinder box plots. The American Statistician 41, 2 (May 1987), 149.



Variations on 2D boxplots - 2

2D Box Plot

- robust line through data
- partition data into 3, find median of outer partitions



P. Tongkumchum, Two-dimensional box plot. Songklanakarin Journal of Science and Technology 27, 4 (2005), 859–866.



Variations on 2D boxplots - 3

Bag plot

- half space depth
- spatial equivalent to quartile statistics



P.J. Rousseeuw, I. Ruts, and J. Tukey. The bagplot: A bivariate boxplot. The American Statistician 53(4), 382–287, 1999.



2D Summary Plot

- Statistics similar to summary plot
- Highlight correlations





Joint Histogram

- Joint Histogram
- Mean & standard deviation of each 1D boxplot





Covariance & Skew Variance

Covariance

- relational variation
- warp ellipse based on covariance matrix

Skew variance

- highlight asymmetries
- arrows scaled by skew variance





Multiple 2D Plots

- Show trends in data
- Covariance and skew variance glyphs distinguish between plots
- 1D summary plots on each side for orientation





User Interface

- User controlled visualization results
- visual clutter reduction
- combine the best plots for specific application





Using the Summary Plots

Short-Range Ensemble Forecasts (SREF)

- Domain across North America
- Forecast weather variables out to 87 hrs
- 4 models using various perturbation schemes (21 members)

Short-range ensemble forecasting. http://wwwt.emc.ncep.noaa.gov/mmb/SREF/SREF.html.







Variance Based Clusters





Standard Deviation



Summary Plots on Clusters

Temperature at 2M above ground, 03/03/2009, Valid Hour 27





Conclusion

- Interactively explore data distributions
- Highlight salient features
 Future Work
- More sophisticated statistics
- Higher spatial dimension
- Parameter investigations
 Coming Soon!! R Package (June 2010)
 http://www.sci.utah.edu/~kpotter/software/kpplots/
 (available now, but very very alpha)

Thanks!

This work was funded in part by

- DOE SciDAC Visualization and Analytics Center for Enabling Technologies (www.vacet.org),
- * NIH NCRR Grant No. 5P41RR012553-10
- * and KAUST (KUS-C1-016-04)