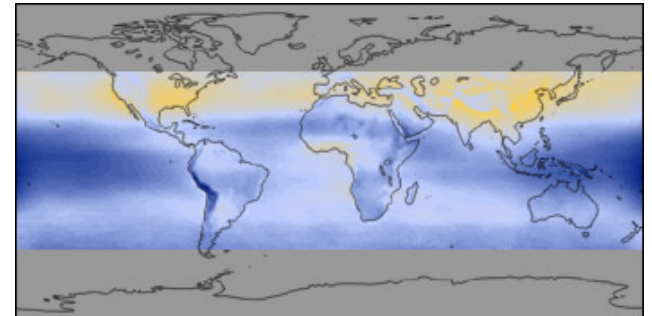


Data Expo 2006

Sponsored by the Sections on
Statistical Graphics, Statistical Computing,
and Statistics and the Environment

August 10, 2005



The data set:

The data are geographic and atmospheric measures on a very coarse 24 by 24 grid covering Central America. The variables are: elevation, temperature (surface and air), ozone, air pressure, and cloud cover (low, mid, and high). With the exception of elevation, all variables are monthly averages, with observations for Jan 1995 to Dec 2000. These data were obtained from the NASA Langley Research Center Atmospheric Sciences Data Center (with permission; see important copyright terms below).

More details about the data, including descriptions of the variables, are available online at www.amstat-online.org/sections/graphics/dataexpo/2006data.php

Download the data (as a zipped tar ball) from there.

The question:

The aim of the Data Expo is to provide a graphical summary of important features of the data set. This is intentionally vague in order to allow different entries to focus on different aspects of the data. For example, the focus can be on: the fact that the data are multi-variate, or time-series, or spatial; or the fact that the data contain missing values; or the focus could even be on the process of exploring the data.

Some obvious general questions that could be answered are: What are the important relationships between the variables? Are there any important trends in the data? Are there any important groupings or clusters in the data? Are there any unusual locations or time periods in the data set?

How to Enter:

Submit a poster abstract for JSM 2006 (all Data Expo abstracts will be included in a special topic poster session; details to be announced).

Student entries and/or group entries are welcome. There is a single competition so student entries may wish to include a staff advisor. Educators may want to incorporate this competition as a class project.

The use of dynamic and/or interactive graphics is likely to be very useful, at least in the exploration of the data. This is encouraged, and we will attempt to provide support for laptops within the poster session so that dynamic/interactive graphics can be included in the poster presentation.

If you are thinking about entering, please send us a quick note so that we can keep you up to date with any updates in the data set or competition rules.

The Prize:

The best posters (as judged by a panel of statistical graphics experts) will be presented with a certificate and a set of NASA books and posters at the Statistical Graphics and Statistical Computing Sections Mixer at JSM 2006. The winner will also receive US\$1000.

Important Dates:

Aug 10 2005: Data released.

Sep 30 2005: Data finalized.

Feb 1 2006: Final date for submission of JSM poster abstracts (online; link TBA). Last day to enter the competition! (but you have until August 2006 to complete your submission)

Other Resources:

NASA provides a large number of images online and free of copyright which may be useful for decorating your poster (see, e.g., Great Images In NASA).

Acknowledgements:

We thank the NASA Langley Research Center Atmospheric Sciences Data Center for providing the data and background information, and for donating the books and posters for prizes. Thanks in particular to Kathleen at ASDC User Services.

NASA copyright:

NASA have allowed us to use their data with the understanding that it will be used for this competition, and request that they receive a copy of all entries so that they have the opportunity to use the material as examples on the MY NASA DATA web site.

Contacts:

For more information and especially for updates to the rules, prizes, and data, visit the Data Expo web site at www.amstat-online.org/sections/dataexpo/2006.php.

Please direct any questions about the data set or the competition in the first instance to Paul Murrell (paul@stat.auckland.ac.nz).