Glaciers Melt as Mountains Warm: A Graphical Case Study

J. Hobbs, H. Wickham, H. Hofmann, D. Cook [a] *

[a]Department of Statistics, Iowa State University,
Ames IA 50011, United States

For the 2006 ASA Data Exposition, we created graphics that, in the legacy of John Tukey, tried to "force the unexpected upon us" (Tukey, 1972). The data are geographic and atmospheric measures on a coarse 24 by 24 grid covering Central America, measured monthly over the course of six years from January 1995 to December 2000.

Using conventional static graphics and some less conventional interactive graphics, we are able to find familiar features in the data, such as seasonal patterns, spatial correlations, El Niño events, as well as some more surprising results, some that corroborate current events in the news.

1. Introduction

Data analysis is messy! Real data sets are rarely perfect and may need multiple passes to ensure high quality, and to converge on findings. At every stage, there are many possible next steps and our path was guided by intuition, knowledge of the phenomenon and past experience. Many steps led to dead ends, and most plots ended up in the wastebasket. This may sound disorganized, but even highly improvisational data exploration does have some structure. Our approach has these components:

1. Refining the questions, deriving expectations
2. Organizing the data
3. Plotting the data
4. Modeling
5. Incorporating other data
6. Presenting findings

The process is not linear, and we returned to re-do, and re-work, the analysis at different stages. Chatfield (1995) has a good description of this process.

The Expo data was daunting! It was complicated by several contexts: spatial, temporal and multivariate. We could have simplified the data by reducing the study to one variable, one time point or one location, but instead we tackled the entire data. Our analysis took many, many hours of labor, re-processing, with dead-ends, lots of plots, discussion, and sharing of code and discoveries.

*This work has been partly supported by the NSF VIGRE Grant0091953.

Rather than simply presenting our findings in the Expo data, this paper also describes the process that led to these discoveries. We think that the process will be interesting reading, both as an example of exploratory data analysis and as a story of the chase for discovery.

2. Getting started

2.1. Refining the questions, deriving expectations

The initial announcement for the Data Expo provided limited general information about the data set and the objectives for analysis. Specifics included the locations and dates of the observations and brief descriptions of the variables. Accompanying the description was a set of four general questions about the data set:

- What are the important relationships between variables?
- Are there important trends?
- Are there important groupings or clusters?
- Any unusual locations or time periods?

Armed with this, we attempted to frame some ideas suggested in these questions into some hypotheses about what we might see in the data:

- The area covers the equatorial Pacific Ocean. We should see El Niño and La Niña effects on the temperature variables, and perhaps other variables, for events that occurred between 1995 and 2000.
- Data from the same month will show correlations between years (seasonal patterns).
- Seasonal trends will be different for sea versus land areas and the northern hemisphere versus the southern hemisphere.
- Data from neighboring locations will be correlated.

To further refine these hypotheses and to develop others, we investigated the way the variables were defined in more detail than was given. We also examined what different climate terms such as El Niño mean.

2.2. Investigating the data sources

From the variable descriptions, it appeared that the observations came from at least two satellite products archived by the International Satellite Cloud Climatology Project (ISCCP). The pressure, ozone and near-surface air temperature variables were derived from the TIROS Operational Vertical Sounder (TOVS). The remaining variables, cloud cover and temperature from clear-sky composite, were likely derived from the Gridded Cloud Product (Rossow and Schiffer, 1991).

The radiation intensity, or brightness temperature, T_B, observed by a satellite is a function of the surface temperature T_{sfc}, the vertical temperature profile in the atmosphere $T(z')$ and some constant quantities. The relationship is described by the basic equation of radiometry (e.g. Ulaby et al. (1981)):

$$T_B = \epsilon T_{sfc} \epsilon^{-\tau(0,\infty)} + \int_0^z \kappa_a T(z') \epsilon^{-\tau(z',z)} dz',$$

where τ is the optical depth, κ_a is the absorption coefficient, z is the height at the top of the atmosphere, and ϵ is the emissivity.

TOVS is calibrated to give a vertical profile of a number of variables, including temperature and ozone. The Gridded Cloud Product is calibrated to distinguish cloud properties and surface characteristics in the absence of clouds. Thus, the two products can give different pictures of surface temperature. In addition, values of the constants can change with changing land surface characteristics, creating several possible sources of error.

2.3. Organizing the data

The data were provided as a set of individual files, one file for each of the variables for a single month—72 files for each variable—each one formatted as follows:

```
VARIABLE : Mean low cloud amount (%)
FILENAME : ISCCPMonthly_avg.nc
FILEPATH : /usr/local/fer_data/data/
SUBSET   : 24 by 24 points (LONGITUDE-LATITUDE)
TIME     : 16-JAN-1995 00:00

           113.8W 111.2W 108.8W 106.2W 103.8W 101.2W 98.8W ...
              27     28     29     30     31     32     33
36.2N / 51:   7.50   7.00   7.00   7.00  11.00  14.50  25.50  ...
33.8N / 50:  11.50  11.50   9.50   8.50  12.50  17.50  27.50  ...
...
```

The four of us came up with different ways to organize the data, mainly triggered by different ways each of us was thinking about the data, emphasizing the spatial, the temporal or the multivariate aspect. The data organization heavily influenced our choices of methods to analyze the data.

Both of our main approaches are in rectangular form, which we found to be the most convenient overall. Even though `R` does allow the creation of data cubes—with two dimensions for the spatial components and one dimension for the temporal component of each variable—it is not easy to use the data in this form afterwards.

The two rectangular formats that we ended up using most often can be described as a "short and wide" version of the data set and a "long and thin" version.

In "wide and short," the focus is on the locations. It therefore has 576 observations, one for each location, and all the other aspects, such as elevation and all monthly measurements of ozone, temperatures, and cloud cover, are encoded in columns:

```
Longitude  Latitude   chJan95   chFeb95   chMar95   chApr95 ...
-113.75    -21.25     0.5       1         2         4       ...
-113.75    -18.75     2.5       0.5       1         0       ...
...
```

That is, each of the original files corresponds to one column of this data set. This version of the data set allows the use of parallel coordinate plots to emulate time series for each location, as seen in Figures 2, 3, 4, 6 and 7.

"Long and thin" has 72 × 576 rows, i.eėach location is repeated for every time index (all 72 months). Measurements such as elevation, which is only recorded once, are repeated accordingly to match the other variables:

```
x y Lat   Long    elev    time Year Month ts    tsa_tovs ...
1 1 36.25 -113.75 1526.25 1    1995 1     272.7 272.1    ...
1 2 33.75 -113.75 612.94  1    1995 1     279.5 282.2    ...
```

Note that a new variable called `time` was created to index time from January 1995 to December 2000. This is then used as the x-position to plot time series. This version of the data allows us to study the multivariate aspects of the data and is the supporting structure of results in Figures 5, 8 and 9.

To keep the geographic context of the data in the forefront, we pieced together a map of the region from Google maps and pinned this to the wall. It is shown in Figure 1.

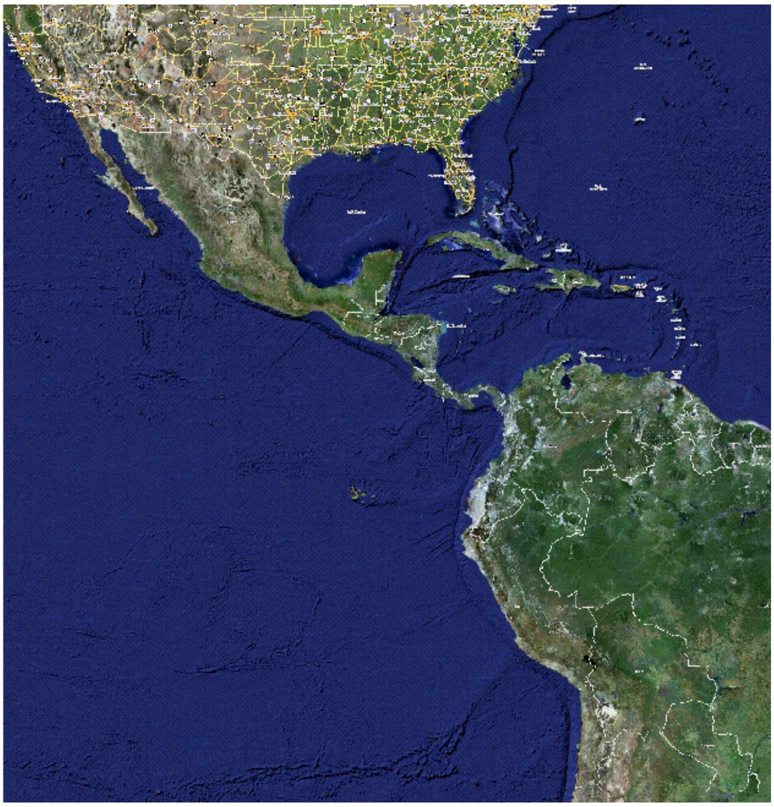

Figure 1. Map of the geographic area where the Expo data was measured, created from Google maps.

3. Plotting the data

3.1. High altitude temporal anomalies

Data exploration often uncovers problems with the data, and this data set is no exception. A look at time series of pressure for each location reveals a problem: between May and June of 1998, several locations exhibit a dramatic increase in pressure, with a number of locations showing an increase of over 100 millibars. Holding all other variables constant, this would be equivalent to a change in elevation of over 1 kilometer! Using the linked graphics in Figure 2, we found that all of the problematic locations were at high elevations.

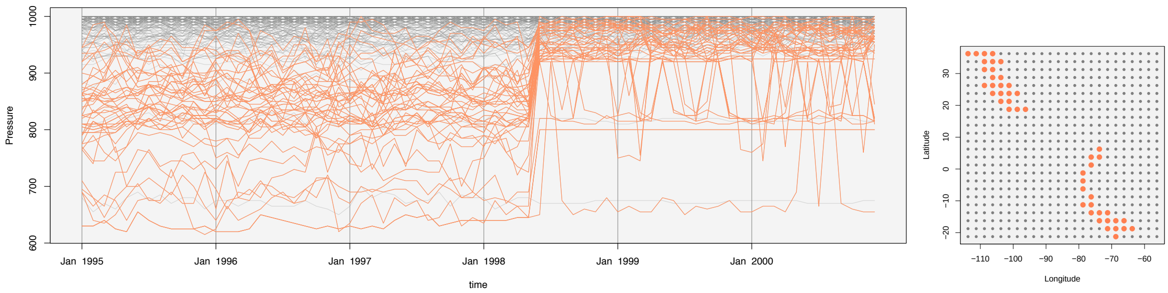

Figure 2. Time series of pressure values for each location. Locations with problematic pressure changes (change of more than 80 millibars) are highlighted in the time series plot and the scatterplot of latitude and longitude. Problematic locations turn out to be in high-altitude areas of the Rockies and Andes.

Looking at temperature in a similar way also revealed several unusual locations. The parallel coordinate plot in Figure 3 shows the temperature profile of all monthly averages, with location 36.25°N, 103.75°W highlighted. After June of 1998 the temperature profile is conspicuously regular; closer inspection reveals that the monthly temperature patterns are repeating in subsequent years. This gave us a strong reason to suspect that these air temperature values were imputed and not observed. Another three locations; 18.75°S, 68.75°W; 21.25°S, 66.25°W; and 18.75°S, 66.25°W; show the same pattern. The imputation used is seasonal and while it does a reasonably good job for the first location, the imputed values for location 21.25°S, 66.25°W are very different from the observed temperature values, as shown in Figure 4.

Because we did not have access to the underlying raw data, we could not investigate this problem further, let alone fix it, so for the rest of the analysis we will ignore pressure and will be wary of the temperature measurements.

We can look at the pressure data in another way, by showing small time series of pressure at location while maintaining the spatial structure, as in Figure 5. The figure was created using a manual correlation tour Cook and Buja (1997); Cook et al. (2006), with the time rotated into the horizontal direction and pressure rotated into the vertical

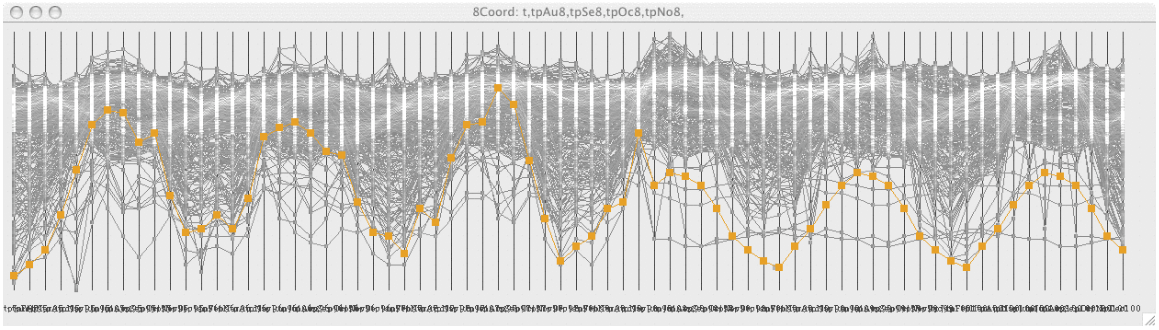

Figure 3. Temperature profiles of all locations. Location 36.25°N, 103.75°W (in the Rockies) is highlighted. The regular pattern of values after June 1998 are an indication that these values are imputed and not observed.

Figure 4. Temperature profiles of all locations. Highlighted is location 21.25°S, 66.25°W (in the Andes of Peru). After June 1998 temperature values are probably imputed but do not reflect previous temperature values well.

direction. The plots in the figure show the results of the sequence of rotations.

Again we can see that the locations where there is a problem with pressure are all in the high-altitude areas. For most of the study area, pressure values are fairly constant over time, but in the higher altitudes, pressure values jump around in the middle of the time series. We can zoom in on one location by linking this plot to a plot of all the time series and brushing one location (bottom plot). We have highlighted a location where some seasonal pattern is visible for the first three years and then the pressure values jump and stay constant for most of the latter three years.

This approach was used with other variables besides pressure. From these we learned that there is something odd about the relationship between the two temperature variables in the high altitudes, and cloud patterns change in the Pacific during an El Niño event. The video accompanying the poster illustrates these findings.

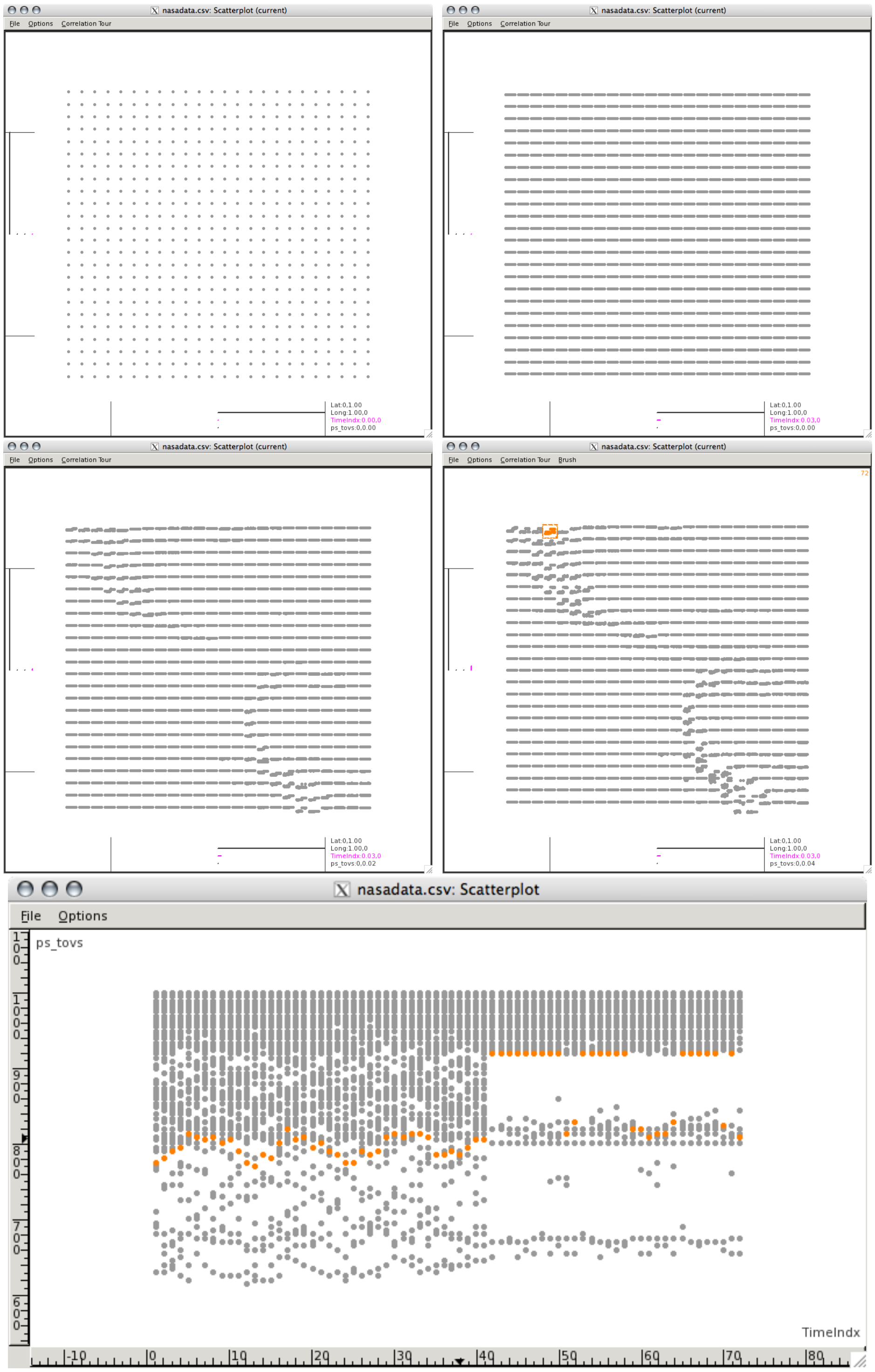

Figure 5. Using a correlation tour to lay out small time series at each location. Pressure values are flat over much of the spatial domain, but in the high altitude areas there is a strange jump or split in values, after the middle of the time series.

3.2. Spatio-temporal trend in ozone

A classical approach to displaying spatial data is to use color to represent the numerical value of a measured variable on a map or geographic area. We used image plots in `R` to plot monthly averages by location. Figure 6 shows the whole set of maps of monthly ozone averages, with color values are scaled across all maps, enabling us to make comparisons across all plots as well as detecting patterns within.

The layout in figure 6 makes comparisons between successive months easy (except for the unfortunate break between December and January) and at the same time allows comparisons of ozone levels for the same month in successive years. This identifies the ozone values of March 1997 as elevated with respect to both February and April of the same year as well as the values in March of all the other years.

Looking at individual plots, we can see a spatial trend: high values are found farther from the equator. Reading across a row, we can see a distinct seasonal trend: in the months of June-October, ozone levels increase from the north towards the equator, and to some extent in the south also. Finally, reading down columns, we can see differences between years: in 2000 the ozone levels have dropped again by September, while in 1995 they stay high through October.

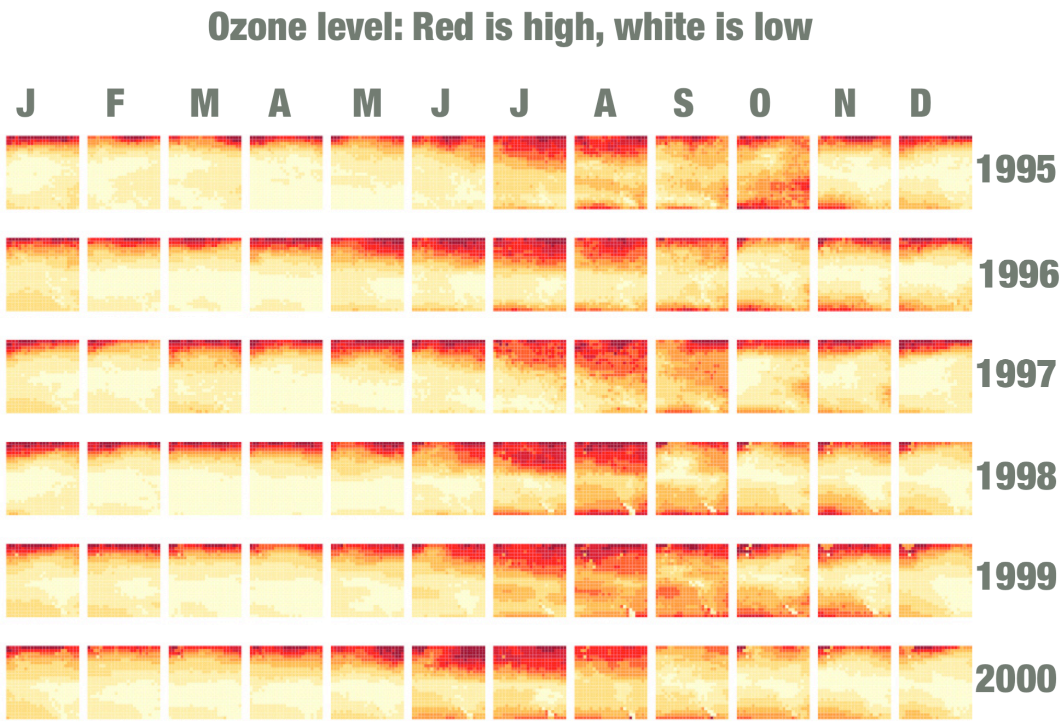

Figure 6. A classical approach to plotting spatio-temporal data: ozone values shown using color on the spatial coordinates, with separate plots for each date.

3.3. The El Niño effect

In late 1997 and early 1998, sea surface temperatures in the equatorial Pacific remained unseasonably high. We can see this in the linked time series of near-surface temperature and spatial location plots in Figure 7. This phenomenon is known as El Niño, and is reported to occur every 2-7 years.

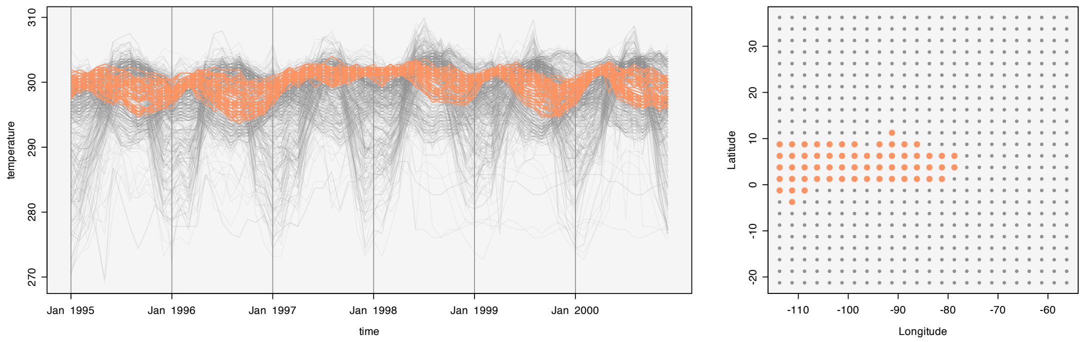

Figure 7. For the location marked in the Pacific Ocean in the scatterplot to the right the temperatures stay unseasonably high during the Winter of 1997 and Spring 1998. This phenomenon is known as El Niño.

Figure 8 explores temporal trends in two locations: one at a northern latitude and another in the equatorial Pacific. We brush in the spatial domain and examine the changes in the time series. We see that at more northern latitudes there is a stronger seasonal trend in temperature and ozone. Low cloud cover exhibits more seasonal trend near the equator, which is noticeably absent during the 1997-8 El Niño event. Pressure is flat in both regions, as expected for low elevation data.

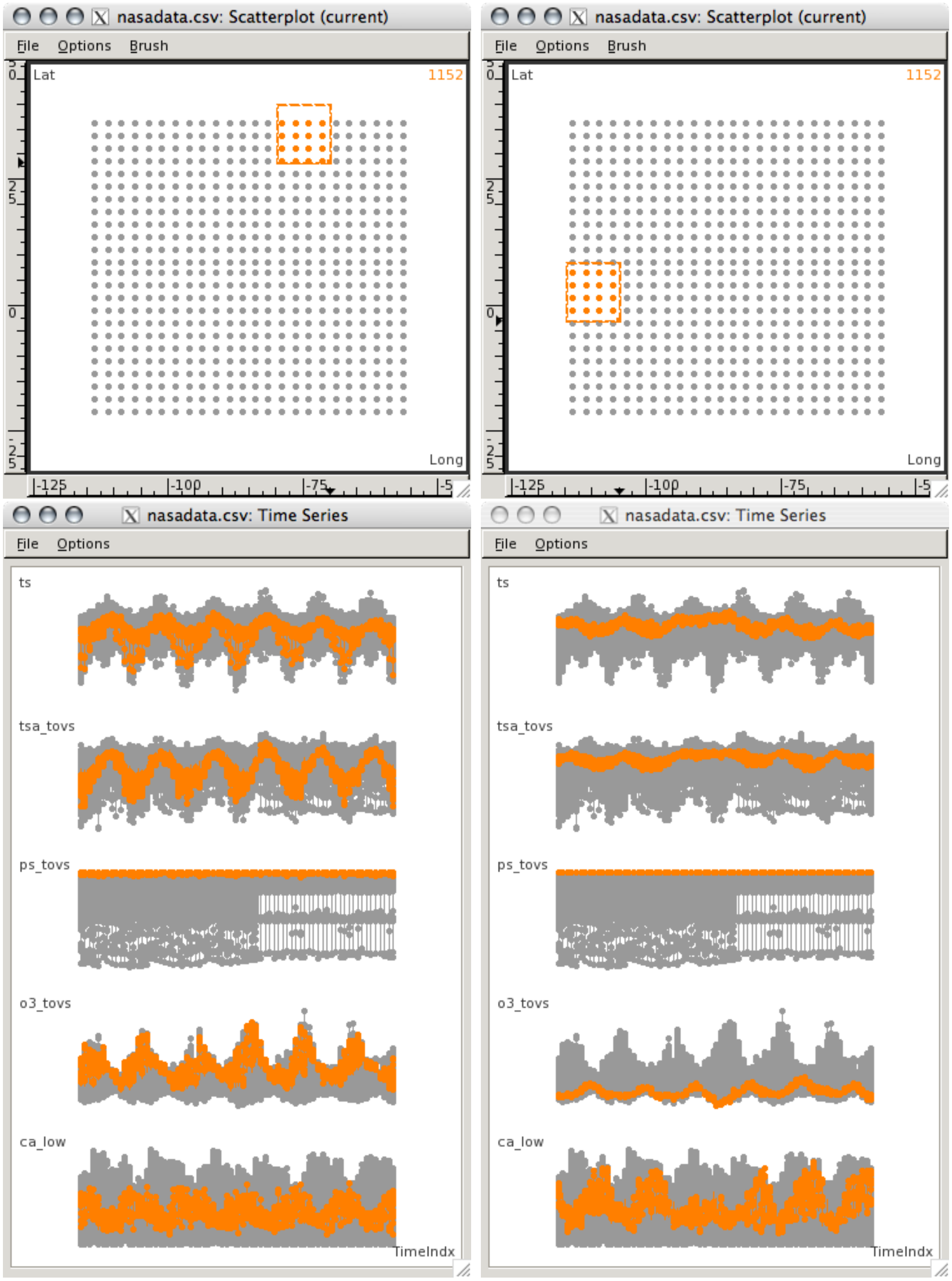

Figure 8. Two views of the data, with linked brushing: one of the geographic locations, and the other of multiple time series of the measured variables. At more northern latitudes there is more seasonal trend in temperature and ozone. Low cloud cover exhibits more seasonal trend near the equator, and it is noticeably absent during the 1997-8 El Niño event. Pressure is flat in both regions, as expected.

3.4. Multivariate relationships

Figure 9 and Table 1 help us examine the pairwise relationships between variables. The two temperature variables are fairly strongly associated. There are a couple of points where there is a big difference between the two, and these are brushed to examine in other views. Further exploration reveals that the largest differences occur on the edges of the spatial region so are probably errors created from the original data processing.

It is more difficult to describe the relationships between other pairs of variables. Pressure has little association with temperature or ozone. Temperature has some association with ozone. The clouds have something of a constrained relationship, for example if there is a lot of high cloud cover there tends to be little low cloud cover. To learn more, we need to condition on time points or locations or both.

Figure 9. Examining all pairwise relationships using a scatterplot matrix of measured variables.

Table 1
Correlation matrix for the measured variables.

	ts	tsa	ps	o3	ca_low	ca_med	ca_high
ts	1.00	**0.81**	0.35	-0.21	-0.07	-0.31	0.04
tsa	**0.81**	1.00	**0.50**	-0.35	-0.20	-0.09	0.22
ps	0.35	**0.50**	1.00	-0.06	0.30	-0.22	-0.09
o3	-0.21	-0.35	-0.06	1.00	-0.01	-0.03	-0.08
ca_low	-0.07	-0.20	0.30	-0.01	1.00	-0.42	**-0.54**
ca_med	-0.31	-0.09	-0.22	-0.03	-0.42	1.00	**0.62**
ca_high	0.04	0.22	-0.09	-0.08	**-0.54**	**0.62**	1.00

4. Modeling

Once we have identified strong patterns, it is useful to remove them and see what remains. In this section we model some of the gross patterns and see what subtle features are left.

4.1. Removing seasonal trends

To focus on long-term trends, it is necessary to remove the strong seasonal pattern. We fit a model to each location; with an intercept, a trend in time and a combination of sine and cosine fits to accommodate for seasonality.

$$Y = \mu + \alpha t + \beta_1 \sin(m) + \beta_2 \cos(m) + \varepsilon, \quad \varepsilon \sim MVN(0, \sigma^2 I)$$

where	Y	temperature (in K)
	t	index of time, $t = 1, ..., 72$
	m	index of month, $m = 2i\pi/12$, $i = 1, ..., 12$
	μ	average temperature
	α	monthly increase in average temperature (in K)
	β_1, β_2	amplitude of temperature
	$(\alpha, \beta_1, \beta_2) \sim N(0, \Sigma)$	

The overall fit is good, $R^2 = 0.91$ and $\hat{\sigma} = 1.47$.

The plot in Figure 10 shows the time trend models as lines and residuals from the fit as points, fitted for each location to the near-surface temperature. Many locations in South America exhibit an increasing trend over time, with some locations showing increases of several degrees per year! Many locations in the equatorial Pacific exhibit a “bump” in the residual pattern late 1997 and early 1998. This anomaly is the El Niño event.

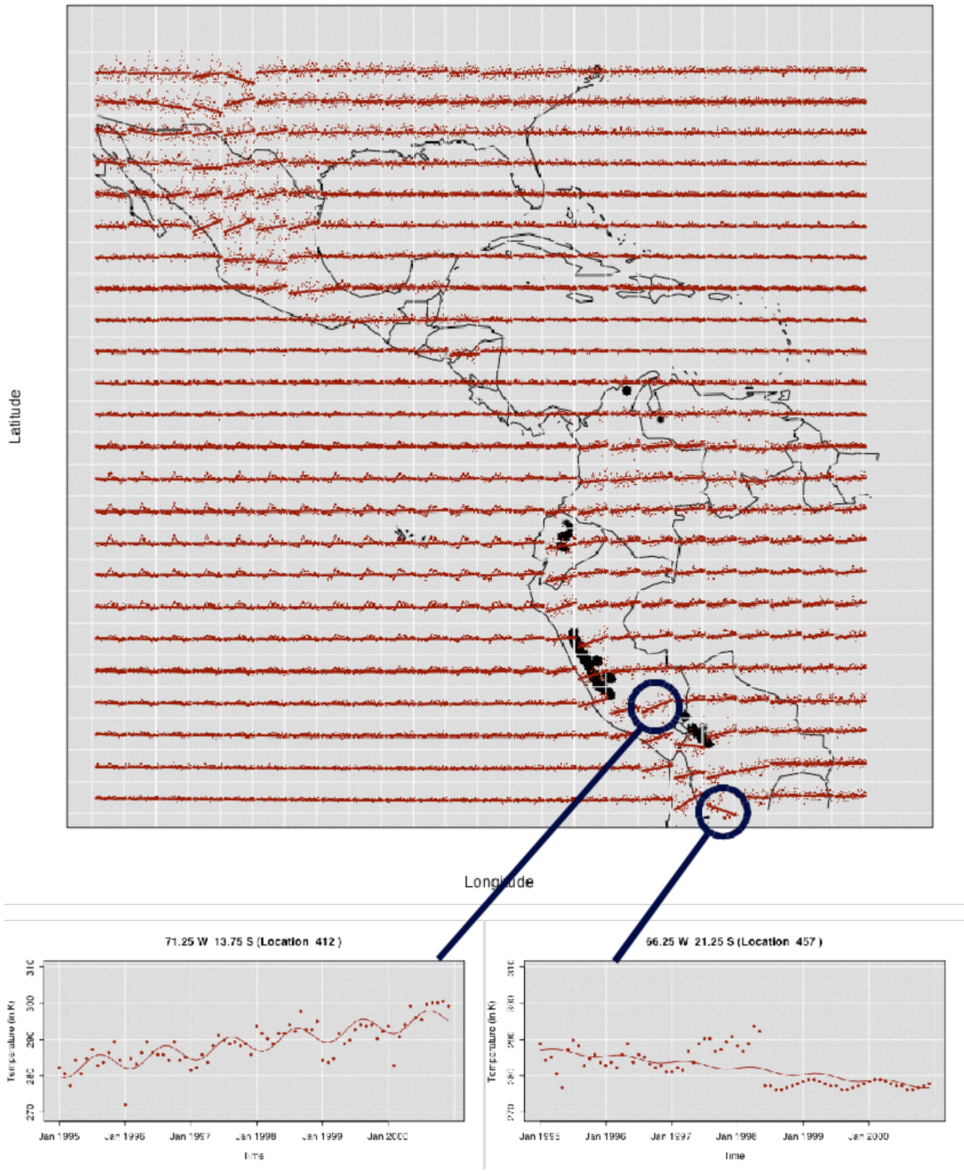

Figure 10. Time trends (lines) and residuals (points) of de-seasonalized temperature values are shown for each location. Large increases in temperature are associated with high altitude. Some of the highest increases, particularly in South America, are in locations of glaciers (marked by black points). The scatterplots below show the temperature values (points) and fitted values (line) for two individual locations. The scatterplot on the right shows one of the locations flagged earlier in section 3.1. The decrease in temperature is artificially introduced by the imputation.

4.2. Clustering ozone

Section 3.2 looked at spatial patterns in ozone, conditioned on time. Similarly, it is worthwhile to look at temporal patterns, conditioned on location. The small multiples in the software Gauguin (Gribov et al., 2006) provide one way of doing this, shown in Figure 11. (This analysis used the wide short form of the data.) At each location, a star glyph displays the amount of ozone measured every month. This graphic allows us to compare patterns at both global and local scales.

Different shapes represent different patterns at a given location: large circular icons indicate high values at all time points; flower shapes indicate seasonal variability, with peaks and dips at the same time each year. Colors categorize the locations into 10 different temporal patterns. These patterns were created using hierarchical clustering.

Global patterns can be seen by looking at the glyphs en masse; try squinting your eyes or looking at the figure from a distance. Ozone levels are highest furthest from the equator and have less seasonal variability in the Northern Hemisphere.

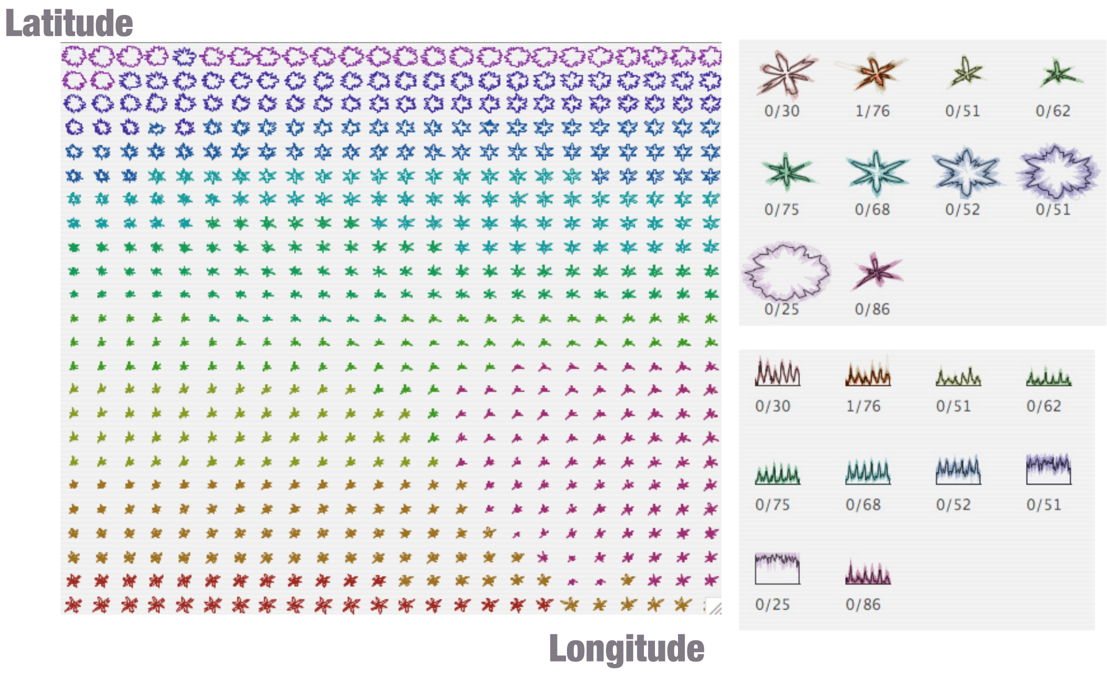

Figure 11. Star glyphs of monthly ozone values between 1995 and 2000. Left: the glyphs are placed according to latitude and longitude. Right: star and profile glyphs showing the 10 different clusters.

4.3. Clustering into local climates

Cluster analysis can help us learn about local climates in the larger geographic region. Ideally, the clustering is done using all of the measured variables, not only ozone. We found that, although Gauguin can theoretically do this, it was too slow. Instead, we used R to cluster the data using all of the variables.

A hierarchical cluster analysis was performed on all variables other than pressure, using the "short and wide" form of the data. Variables were standardized to have zero mean and unit variance. Interpoint distance was measured by Euclidean distance and Ward's linkage was used to describe distances between clusters. Ten clusters were chosen based on the dendrogram and from our own limited knowledge about the geographical regions. The cluster classification of each location is shown in Figure 12. The measurements in each cluster are shown as time series plots in the right-hand side plot.

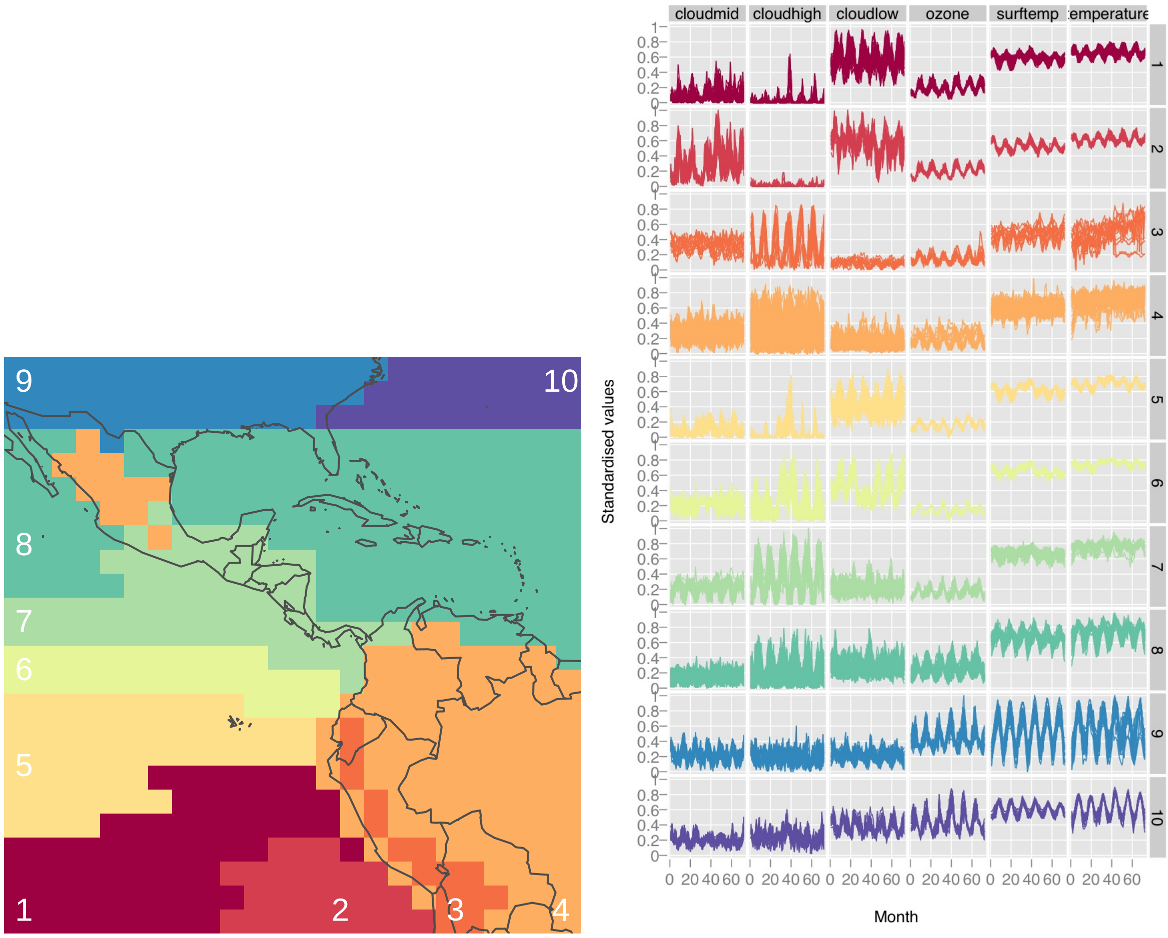

Figure 12. Hierarchical clustering using Ward's method, with (left) geographic location of clusters and (right) time series of variables showing differences between clusters.

Clusters 1 and 2 have highly varied levels of mid-level and high cloud cover and small seasonal trends in ozone and temperature. Cluster 1 has a big peak in high cloud cover during the middle of the time period, the El Niño event. This peak is evident in cluster

5, too. Cluster 2 has very low levels of high cloud cover. Cluster 9 in the north and cluster 10 over the South American coastline have the most seasonal variability in ozone. Cluster 7, which includes most of South America and the central part of Mexico, shows increasing temperatures. Although we could spend a lot of time digesting the intricacies of these clusters, the main message to be learned is that the clusters correspond to contiguous geographic regions, roughly matching what we know about the geography: Pacific regions, Caribbean, land/sea. We also might expect that we have artificially partitioned a continuous gradient into discrete chunks.

5. Incorporating other data

5.1. Where are the glaciers?

We researched the locations of glaciers in South America and overlaid these in black on the map view in Figure 10, back several pages. The most dramatic increases in temperature occur at high elevations and near the locations of glaciers.

Searching for more information on glaciers and warming on the internet also uncovered the photos of the Qori Kalis glacier in Peru taken in 1978 and 2000, which show that this glacier has retreated substantially over two decades. These same photos were used in the 2006 movie, “An Inconvenient Truth”.

5.2. Is the satellite data verified by ground stations?

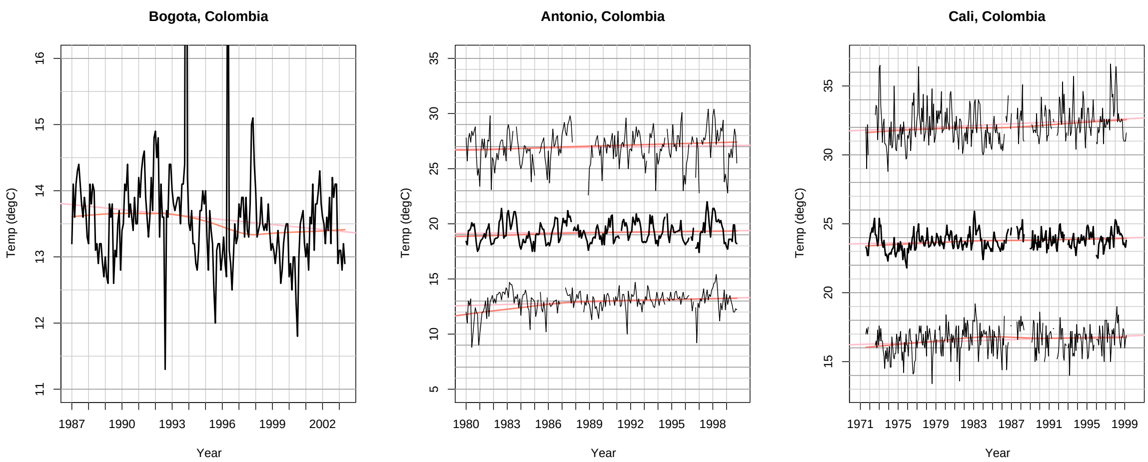

Figure 13. Monthly temperature data from recording stations in Colombia: Bogota (4.43°N, 74.9°W, Elev. 2548m), Antonio (1.25°N, 77.16°W, Elev. 1826m), Cali (3.33°N, 76.23°W, Elev. 969m). Minimum, mean and maximum are shown for Antonio and Cali, mean only for Bogota. Loess curves are overlaid on the data.

To investigate the apparent increase in temperatures at high altitudes we searched for information from ground-recording stations in the vicinity. The closest we could find are in Colombia, near Bogota, Cali and Antonio. (There are glaciers in Colombia which are also reported to be receding.) The temperature data are shown in Figure 13. The measurements cover varied time frames, starting earlier than the Expo time period. In each case a loess curve is fit to the data. Measurements for Bogota are erratic, but values for Antonio and Cali, where we have minimum, maximum and average values, indicate some increasing trend over the time period, although not as strong as the Expo data.

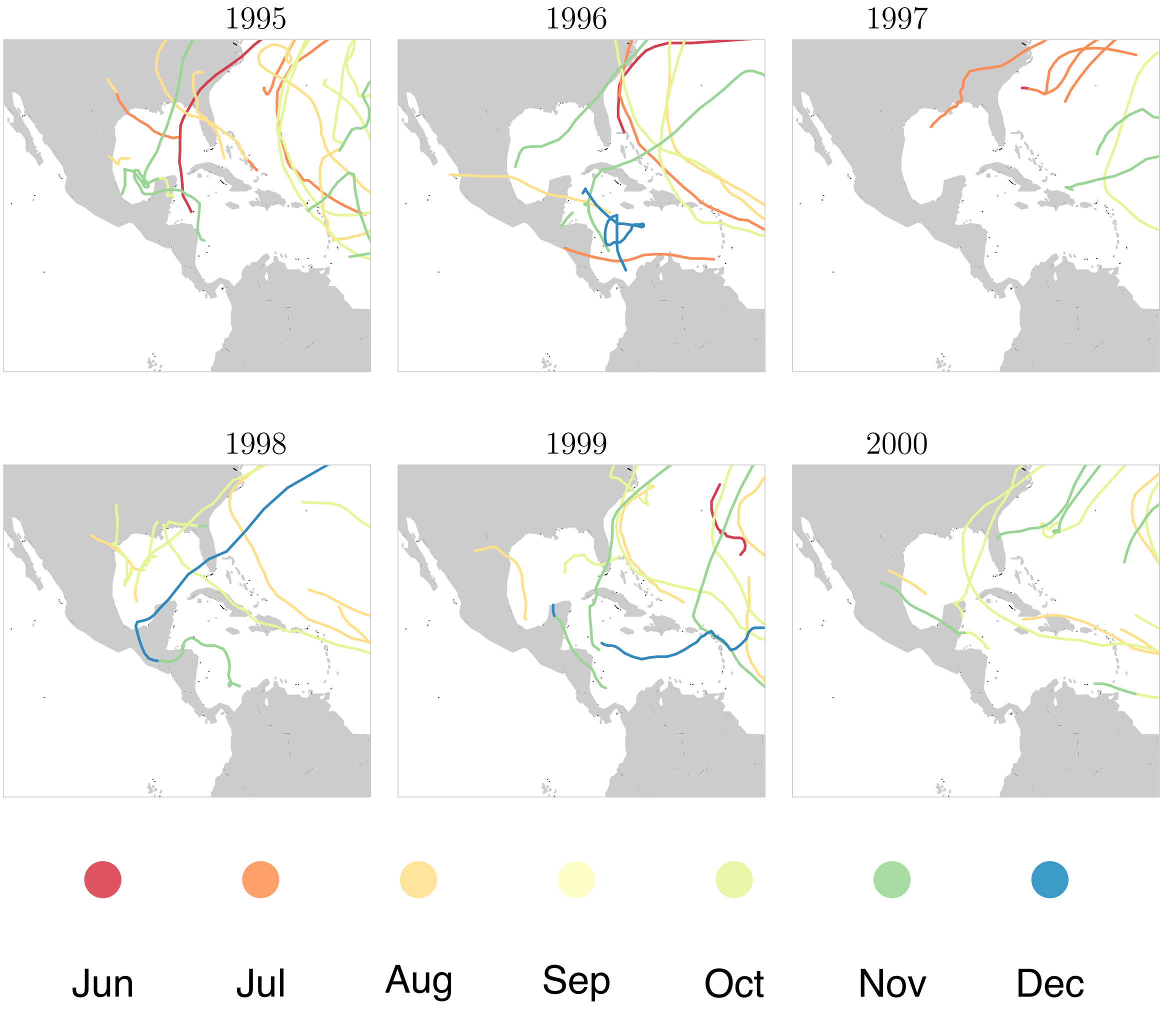

Figure 14. Hurricane activity seems to slow down during the time of the El Niño in Fall 1997 to Summer 1998. There are early storms in 1997, but few late storms, and mostly late storms in 1998.

5.3. How are El Niño and tropical cyclones related?

To investigate how El Niño and tropical cyclones are related, we constructed tropical storm track data from the National Hurricane Center's (NHC) archive of tropical cyclone reports for the Atlantic, Caribbean and Gulf of Mexico. The best track storm data record the latitude/longitude position of each storm's center and the surface air pressure at the center every six hours during the storm's lifetime. Additionally, the maximum sustained wind speed and the storm's stage are included.

Information in the tropical cyclone reports was compiled by many different analysts at the NHC, and this was especially evident in the description of the storm stage. For our data set, the stages were classified into one of four categories: Hurricane, Tropical Storm, Tropical Depression or Extratropical (an individual storm could and often did go through all four stages). Only named storms (those that had Tropical Storm or Hurricane status at some point) were included.

Tropical cyclones are among the most noteworthy weather events that impact the Western Hemisphere. With this tropical cyclone data, we attempted to link the location, timing, frequency and length of tropical cyclones to other findings in the original data. The El Niño of 1997-98 was evident in some of the variables in the Expo data (temperature, cloud cover) and El Niño has been linked to climate anomalies all over the world.

In this case, the storm tracks revealed a remarkable absence of hurricane activity in the Atlantic in late 1997 and mid 1998, as shown in Figure 14. This is one illustration of the overall findings of Bove et al. (1998), who investigated the historical frequency of tropical cyclones in the Atlantic basin. They concluded that activity is reduced during El Niño events.

6. Summary of discoveries

These are the answers to our initial questions:

- We can see the El Niño event strikingly in time series plots of the temperature.
- All variables show seasonal trend in some locations, mostly in the higher and lower latitudes: the same month is correlated between different years. The exception is the El Niño event, 1997-1998, in the equatorial Pacific, where temperatures and cloud patterns do not return to similar values from the previous year.
- There are clearly differences between land and sea. Even though we did not focus on these differences in our analysis, the difference between land and sea comes out automatically in the cluster analysis and has about as much effect as the differences in latitude.
- The variables all exhibit spatial dependence: neighboring areas have similar values for the most part. There are some large jumps going from land to sea and also in the high altitude areas where a small spatial difference corresponds to a big difference in elevation.

Most of our time was spent on investigating the surprising discoveries in more detail. We actually learned some things about climate change, that were only rumors to us prior to this analysis. Each of us is more attuned to the stories of glacial melt, and fracturing of ice shelves, as a consequence of looking at this data. Here is a summary of our unexpected findings:

- Temperatures at high altitudes have been increasing over the time period; in some locations as much as several degrees per year, instead of following the usual seasonal peak and decline. These locations match locations of glaciers.

- There is obviously a problem with the pressure variable, so for the most part we left this variable out of our analyses.

- Storm tracks data in the Atlantic basin indicate a decline in the number and severity of storms near the one El Niño event in the study period. We might be able to follow up on this by summer 2007 following the (weak) El Niño of 2006.

- The patterns in cloud cover changes during the El Niño event. In the equatorial Pacific, the high cloud peaks during the event, as does low cloud, and medium cloud is lower than normal.

Ideally we would get some feedback from the data experts at NASA about these findings.

7. Tools and additional data sources

If you are interested in seeing our 4' x 8' poster, or watching the movies we presented at the JSM, you can visit our data expo website: `http://had.co.nz/dataexpo`.

`R` (R Development Core Team, 2006) and the packages `ggplot` (Wickham, 2006) and `rggobi` (Lawrence and Wickham, 2006), were big players in statistical analysis and producing graphics. MANET (Unwin et al., 1996), GGobi (Swayne et al., 2003), and Gauguin (Gribov et al., 2006) were used for dynamic and interactive graphics.

We maintained working documents describing findings and re-structured data sets in a central storage location so that all of us could access them. Since each of us tended to use different software, we periodically used this exercise as a chance to educate others, giving software tutorials on the Expo data exploration for each other and to colleagues in the department.

We supplemented the data provided by the organizers with additional data sourced over the web:

- Positions of glaciers are provided in form of the World glacier inventory by the National Snow and Ice Data Center at Boulder, CO.
  `http://nsidc.org/data/glacier_inventory/`

- Qori Kalis glacier photos, 1978 and 2000
  `http://researchnews.osu.edu/archive/andespics.htm`

- Temperature data from recording stations in Colombia are available from the National Climatic Data Center at the National Oceanic and Atmospheric Administration (NOAA),
  `http://gis.ncdc.noaa.gov/website/ims-cdo/gsom/viewer.htm`

- Storm track data were provided by the NOAA National Hurricane Center at
  `http://www.nhc.noaa.gov/pastall.shtml`

- ISCCP information ISCPP is the International Satellite Cloud Climatology Project, as part of the World Climtae Research Program and can be reached at `http://isccp.giss.nasa.gov/`

References

Bove, M. C., O'Brien, J. J., Eisner, J. B., Landsea, C. W., and Niu, X. (1998), "Effect of El Niño on U.S. Landfalling Hurricanes, Revisited," *Bulletin of the American Meteorological Society*, 79, 2477–2482.

Chatfield, C. (1995), *Problem Solving: A Statistician's Guide*, `http://www.crc.com`: Chapman and Hall.

Cook, D. and Buja, A. (1997), "Manual Controls For High-Dimensional Data Projections," *Journal of Computational and Graphical Statistics*, 6, 464–480, also see `www.public.iastate.edu/∼dicook/research/papers/manip.html`.

Cook, D., Lee, E.-K., Buja, A., and Wickham, H. (2006), "Grand Tours, Projection Pursuit Guided Tours and Manual Controls," in *Handbook of Data Visualization*, `http://www.springer.com`: Springer, p. To appear.

Gribov, A., Unwin, A., and Hofmann, H. (2006), "About Glyphs and Small Multiples: Gauguin and the Expo." *Statistical Computing and Graphics Newsletter*, 17, 14–17.

Lawrence, M. and Wickham, H. (2006), *rggobi: Linking R and GGobi*, R package version 2.1.4.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rossow, W. B. and Schiffer, R. A. (1991), "ISCCP Cloud Data Products," *Bulletin of the American Meteorological Society*, 72, 2–20.

Swayne, D. F., Lang, D. T., Buja, A., and Cook, D. (2003), "GGobi: Evolving from XGobi into an Extensible Framework for Interactive Data Visualization," *Journal of Computational Statistics and Data Analysis*, 43, 423–444, http://authors.elsevier.com/sd/article/S0167947302002864.

Tukey, J. W. (1972), "Exploratory data analysis: as part of a larger whole." in *Proceedings of the 18th Conference on Design of Experiments in Army Research and Development I*, Washington, DC, p. 110.

Ulaby, F. T., Moore, R. K., and Fung, A. K. (1981), *Microwave Remote Sensing: Active and Passive*, vol. 1, Reading, MA: Addison-Wesley.

Unwin, A., Hawkins, G., Hofmann, H., and Siegl, B. (1996), "Interactive Graphics for Data Sets with Missing Values - MANET." *Journal of Computational and Graphical Statistics*, 5, 113–122.

Wickham, H. (2006), *ggplot: An implementation of the Grammar of Graphics in R*, R package version 0.4.0.