Comput Stat (2010) 25:587–601
DOI 10.1007/s00180-010-0210-8

ORIGINAL PAPER

A two-step process for graphically summarizing spatial temporal multivariate data in two dimensions

Svetlana K. Eden · Angel Q. An · Jeffrey Horner · Cathy A. Jenkins · Theresa A. Scott

Received: 16 February 2007 / Accepted: 16 July 2010 / Published online: 29 July 2010

Abstract Graphical illustration and exploration of atmospheric data, like those from the 2006 ASA Data Exposition, is challenging. We chose to employ a two-step process, which consisted of exploring the presented data (1) from a 'global' view by simultaneously displaying the spatial and temporal components of individual variables in the data set using a geographic grid of polar coordinate plots or circular histograms that incorporated elevation as a background color, and (2) from a 'local' view by displaying the relationships between multiple variables at specific geographic locations and/or time points/periods (selected via step 1) using scatterplot matrices.

Keywords Atmospheric data · Graphics · 2006 ASA Data Exposition

1 Introduction

Graphics are the visual means by which statisticians convey their data's content, structure, and analysis results. Often, multiple variables have been collected to give a snapshot in time as a series, representing independent units of measurement. Graphs, such as histograms, boxplots, strip plots, dot plots, and scatter plots, are employed to illustrate the distribution of individual variables and/or relationships between variables. With simple data sets, choosing the appropriate graphs that capture the important aspects of the data is fairly straightforward. On the other hand, the graphic display of data, such as time series or spatial data, presents a complex challenge. In these instances, data no longer consist solely of a multivariate component. Instead, it additionally encompasses a spatial (geographically related locations) and/or temporal

S. K. Eden · A. Q. An · J. Horner · C. A. Jenkins · T. A. Scott (✉)
Department of Biostatistics, Vanderbilt University, 1161 21st Ave. South,
S-2323 MCN, Nashville, TN 37232-2158, USA
e-mail: theresa.scott@vanderbilt.edu

(collected over time) component. In turn, the ability to capture all aspects of the data in specific graphs is not necessarily feasible.

Such is the case with the atmospheric data set presented for use in the 2006 ASA Data Exposition Competition. The data consisted of eight variables: elevation (meters), surface temperature from a clear sky composite (degrees Kelvin), near-surface air temperature (degrees Kelvin), surface pressure (millibars), ozone abundance (Dobson units), low cloud coverage (percent), medium cloud coverage (percent), and high cloud coverage (percent). Each variable was measured at 576 geographic locations based on a very coarse 24 by 24 spatial longitude/latitude grid covering Central America. Furthermore, each variable, except elevation, was measured at each geographic location at 72 points in time, once per month from January 1995 to December 2000. Elevation was measured once in January 1998. Therefore, each variable, except elevation, actually represented a monthly average value for each geographic location.

Considering the complexity of the Data Exposition data and our unfamiliarity with atmospheric data in general, we consulted the current literature to explore the graphical methods often used in meteorology. These methods included time series plots, histograms, contour maps, and dot diagrams. For example, in the ozone literature, Carslaw (2005) used time series plots to describe the shift of ozone maxima, while Lindskog et al. (2003) employed heat maps in addition to time series plots to characterize spatial variability of maximum ozone concentrations. In addition, Lehman et al. (2004) used time series plots to characterize spatial variability of maximum ozone concentrations, Vukovich (1994) used histograms and contour maps to display the seasonal spatial distribution of average daytime maximum ozone concentrations, and Scheel et al. (1997) used time series plots, histograms, and colored 'choropleths' to display average seasonal variations of ozone and spatial distributions of ozone concentrations.

Typically, these types of tools are able to capture one component of a *single* atmospheric variable, either the spatial or temporal; however, they are often unable to capture both components simultaneously. We also have to consider how to capture the relationships between *multiple* atmospheric variables (i.e., the multivariate component of the data). With this in mind, our response to the Data Exposition challenge was to focus on the process itself of graphically exploring the presented data. Specifically, we employed a two-step process, examining the data set

1. from a 'global' view by simultaneously displaying the spatial and temporal components of individual variables using a geographic grid of polar coordinate plots or circular histograms that incorporated a background color to represent elevation, and
2. from a 'local' view by displaying the relationships between multiple variables at specific geographic locations and/or time points/periods, selected via step 1, using scatterplot matrices.

It should be noted that the graphical tools used in the remainder of this paper are not new; a grid of plots, polar coordinate plots, circular histograms, and scatterplot matrices have all been employed to graphically illustrate countless other data sets (Chambers et al. 1983). It is our strategic application of these graphical methods used to explore challenging data, such as the Data Exposition data set, that we feel is a fresh approach. Our intended audience is anyone interested in graphical display who may

or may not be aware of the methodologies discussed in this paper. Lastly, all graphics were created using R (R Development Core Team 2009).

2 Exploring the data from a global view: simultaneously displaying the spatial and temporal components of individual variables

2.1 Displaying the spatial component

As explained earlier, the data covered Central America using a coarse 24 × 24 grid representing 576 distinct geographic locations. At first, we considered creating separate graphs of a single variable at each geographic location. However, finding overall patterns across all the individual graphs proved difficult at best. In order to create a holistic 'global' view of the overall patterns of a single variable with respect to the data's spatial component, we created a matrix or grid of graphs with each graph representing a distinct geographic location. Within each graph, elevation was represented by a background color: blue for sea level at 0 m elevation and a *continuous* white-to-orange color gradient for elevations greater than 0 m, with darker orange implying higher elevations. The continuous white-to-orange color gradient was calculated by converting the negative square root of each unique non-zero elevation to a Red-Green-Blue (RGB) triplet in hexadecimal format (e.g., "#0000FF" represents the color blue). This allowed each unique non-zero elevation to be represented by its own unique color. If desired, this color scheme could be easily converted to grayscale using white for ocean and a continuous grayscale for land, with darker gray implying higher elevations. To further delineate the spatial component, we overlaid a map of the region onto the existing grid of graphs. This grid of graphs with background color denoting elevation is illustrated in Fig. 1. Note, the six-level color scale shown below the x-axis in Fig. 1 illustrates the colors calculated for six specific elevations in the data set (i.e., they do not represent ranges of elevations): 0 m (blue), 0.06 m (minimum non-zero elevation; white), 112.69 m (lightest orange), 1,040.31, 2,231.31, and 4,268.81 m, (maximum elevation; darkest orange).

2.2 Displaying the temporal component

Next, we considered how a single variable's temporal component should be depicted in our grid of graphs. Typically, a single variable's temporal component is displayed using time series, where the values of the variable are plotted on the y-axis and time is plotted on the x-axis. Times series are suitable to display both periodic (repetitive) and long-term features of single variables. However, because each variable in our data set was collected monthly for a relatively short 6 years, we focused on a single variable's periodic features, most specifically its seasonality, rather than possible long-term trends. As an example, Fig. 2 displays the seasonality of ozone at a *single* geographic location using the typical time series plot. Unfortunately, as Fig. 3 illustrates, when this type of graphical display is expanded to incorporate the spatial component of the data set using the grid of graphs discussed in Sect. 2.1, it becomes difficult to distinguish

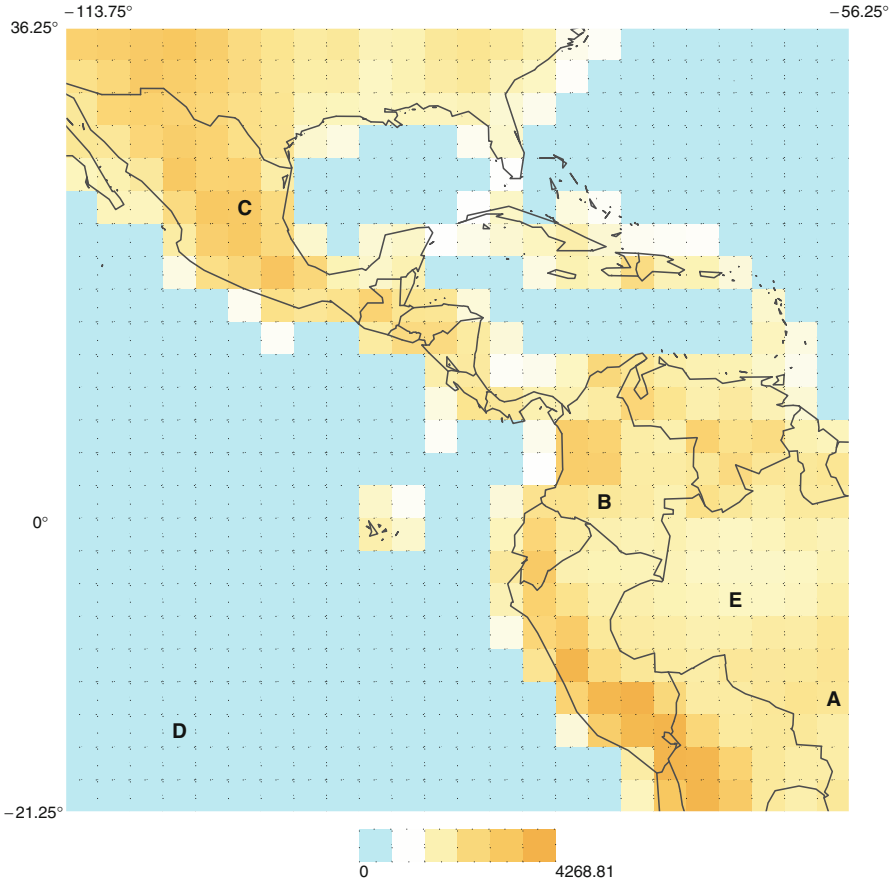

Fig. 1 Grid of graphs depicting the 576 (24 × 24) distinct geographic locations (covering Central America) that are represented in the data set. Each geographic location is delineated by the *dotted lines*. The *background color* of each graph denotes elevation (in meters): *blue* denoting zero elevation (i.e., sea level) and a *continuous white-to-orange* color gradient denoting elevations greater than 0 m, where *darker orange* implies higher elevation. The continuous *white-to-orange* color gradient was calculated by converting the negative square root of each unique elevation to a *Red-Green-Blue* (RGB) triplet in hexadecimal format. The six-level color scale shown below the x-axis illustrates the colors calculated for six specific elevations in the data set (i.e., they do not represent ranges of elevations): 0 m (*blue*), 0.06 m (minumum non-zero elevation; *white*), 112.69 m (*lightest orange*), 1,040.31, 2,231.31, and 4,268.81 m (maximum elevation; *darkest orange*). A map of the region is also overlaid for further distinction of the geographic locations. The locations labeled A through E are used for illustrative purposes in Sect. 3

seasonality, particularly its possible variability from year to year, in great detail both at any single geographic location and across multiple geographic locations.

To address this challenge, we used polar coordinates to plot the temporal component of a single variable at a single geographic location. A polar coordinate system consists of a pole and a polar axis, or a ray, that originates from the pole. A point in the polar coordinate system is defined by two coordinates: the distance from the pole (r) and the angle from the polar axis (alpha). The distance from the

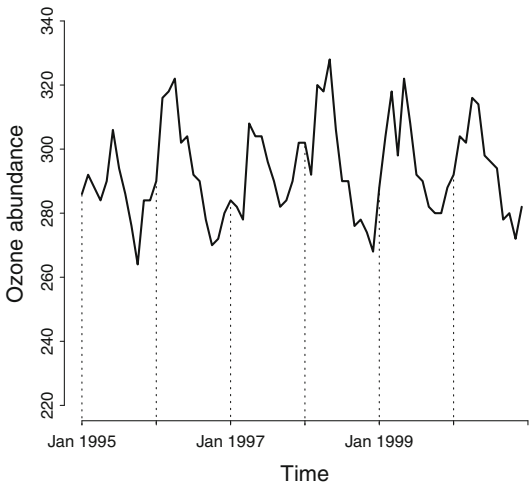

Fig. 2 Traditional time series plot used to display a variable (here ozone abundance) across time

origin, r, is the magnitude of a measure of interest. For example, to display the seasonality of ozone, r would represent the ozone abundance variable in our data set.[1] The angle from the axis, alpha, represents a unit of time (e.g., 1 month). In order to display the seasonality of a single variable and its possible variability from year to year, we projected 1 year's data (12 monthly values) onto one full rotation of alpha (a full circle; 2π radians), such that months were incremented every $\frac{\pi}{6}$ radians with January occurring at 0 radians and December at $\frac{11\pi}{6}$ radians (i.e., angle increases counter-clock wise). In this sense, using consecutive revolutions of the same circle, each successive year's data was plotted on top of the previous year's data.

In Fig. 4, the seasonality of ozone abundance at a single geographic location is plotted using polar coordinates. Figure 5 extends the polar coordinate plot to the grid with line color differentiating sea level (0 m elevation; blue) from elevations greater than 0 m (dark red). Compared to the corresponding time series plot in Fig. 2, the overlapping circles in Fig. 4 more clearly illustrate the variability in the seasonality of ozone abundance across the 6 years of data. In particular, the ozone abundance varies far less in the months of June to September compared to the months of January to June. Although, it is no longer possible to distinguish which circle represents which year of data, it is possible on the grid of graphs to distinguish clusters of geographic locations with similar variability in the seasonality of ozone abundance across all 6 years. For instance, at any specific longitude, ozone abundance from January to June in the northern geographic locations (in or north of Mexico) varies greatly across the 6 years. In contrast, ozone abundance appears fairly stable throughout the year and across all 6 years in the southern geographic locations, especially those in the higher elevations of Central America.

Even though polar coordinate plots were useful in exploring variability in the seasonality of a specific variable, we also wanted to explore how to graphically

[1] Note, all values of r must be non-negative (≥ 0). If the original values of r include negative values, a transformation must be used to generate all non-negative values (e.g., deviation from a 'global' minimum).

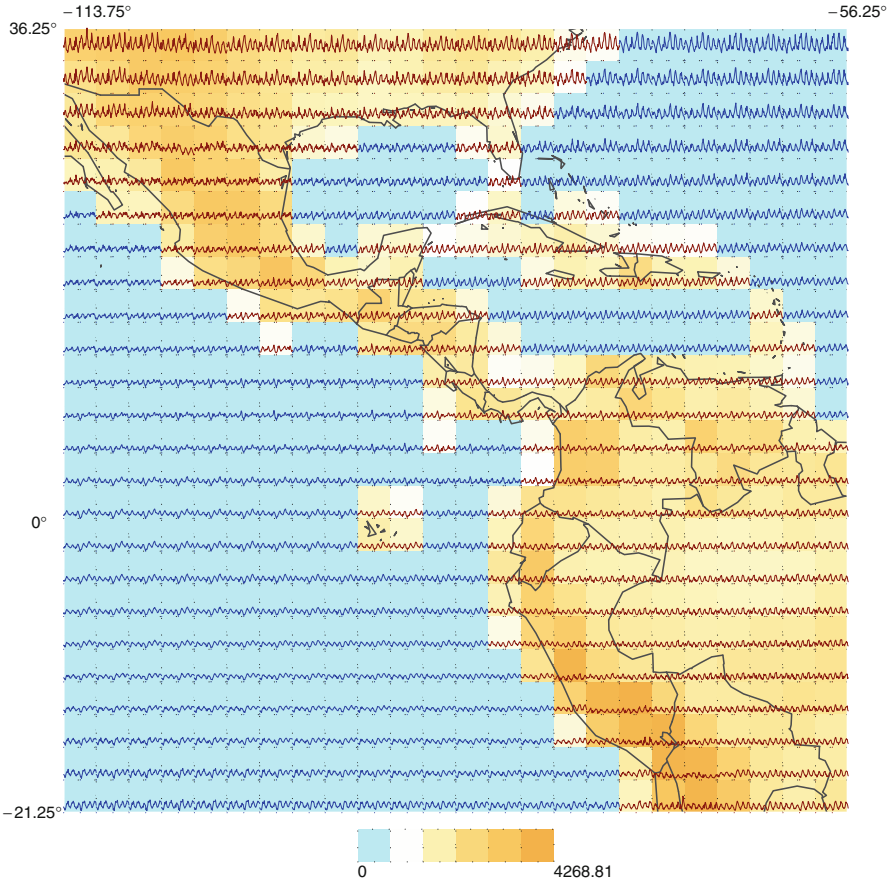

Fig. 3 Ozone abundance time series plotted over a map of Central America overlaid with a 24 × 24 grid of elevation (in meters), with *lighter orange* denoting lower elevation, *darker orange* denoting higher elevation, and *blue* denoting zero elevation. Each geographic location is delineated by the *dotted lines*. The *line color* of the data points is used to further differentiate elevation with *blue* representing zero elevation and *dark red* representing greater than zero elevation

illustrate the frequency of a specific temporal event, such as in what month of each year did a variable reach its maximum? Several circular graphical summaries are available to accomplish this including polar histograms, Nightingale's coxcombs, and circular histograms (Wilkinson 1999). In a polar histogram, the angle represents orientation in space (i.e., 0–360°), not time. A Nightingale's coxcomb traditionally summarizes rates of *multiple* events for a *single* year. Unlike polar histograms and Nightingale's coxcombs, we were interested in summarizing the frequency of a *single* event across *multiple* years. Thus, we turned to circular histograms, which show the timing (season) and persistence (frequency) of an event of interest across multiple years and offer a more focused view of a particular aspect of a variable than that of polar coordinate plots. To plot a circular histogram, the frequency in each year of the event of interest must first be calculated from the available data. Specifically, those months

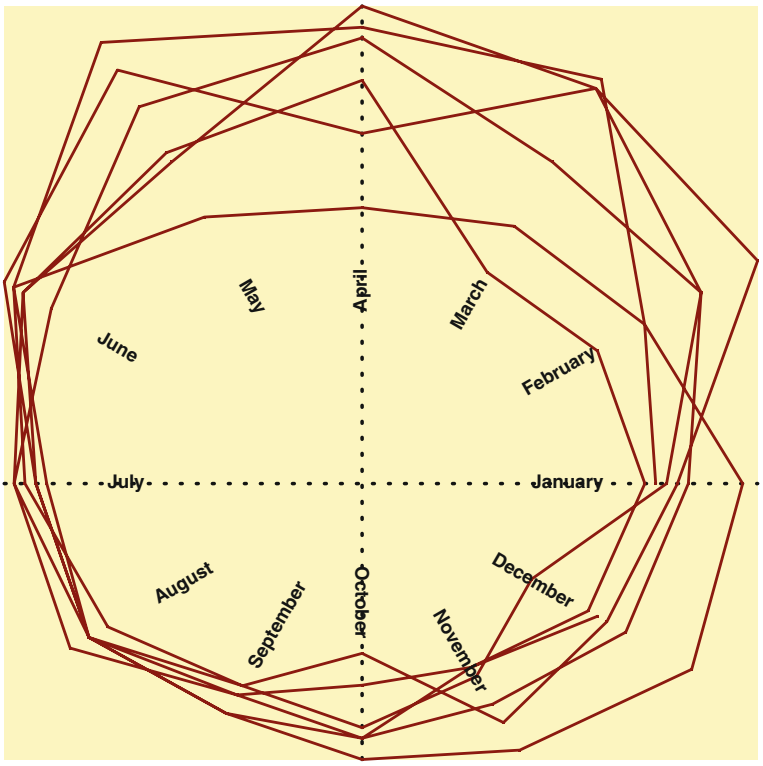

Fig. 4 Ozone abundance time series plotted in polar coordinates. In polar coordinates, each point is defined by a pair, a radius and an angle, where angle increases counter-clock wise. The time-starting point, January, corresponds to zero radians. Each successive month's angle is $\frac{\pi}{6}$ radians larger than the previous one. Each year's time series overlaps with the previous year. Here, the radius is the value of ozone abundance and the angle is time since the beginning of the year

(or time periods) in which the event of interest occurred are assigned a frequency of 1/(number of events of interest in the given year); all other months are assigned a frequency of 0. The monthly frequencies are then summed across all years and divided by the total number of years, resulting in 12 average monthly frequencies defined for each month. A more detailed definition of a circular histogram is given in the Appendix.

As an example, consider 3 years of monthly ozone abundance observations from January 2000 to December 2002. Studying the timing of maximum ozone is very important in environmental research since knowing the seasonal cycle of ozone allows us to ascertain whether ozone comes from un-polluted or polluted air mass (Carslaw 2005; Lehman et al. 2004; Lindskog et al. 2003; Monks 2000; Scheel et al. 1997; Vukovich 1994). The maximum ozone abundance for the 2000 data occurs in March and June; the maximum of 2001 occurs in March; and the maximum of 2002 occurs in April. Because the maximum of 2000 occurs in 2 months, March and June have frequencies of 1/2. In 2001, March has a frequency of 1 since it is the only month in which the maximum occurs. Likewise in 2002, April has a frequency of 1. Summing across the 3 years, the average monthly frequencies of March, April, and June are 1/2, 1/3, and 1/6, respectively. This example is illustrated in Fig. 6. The angle

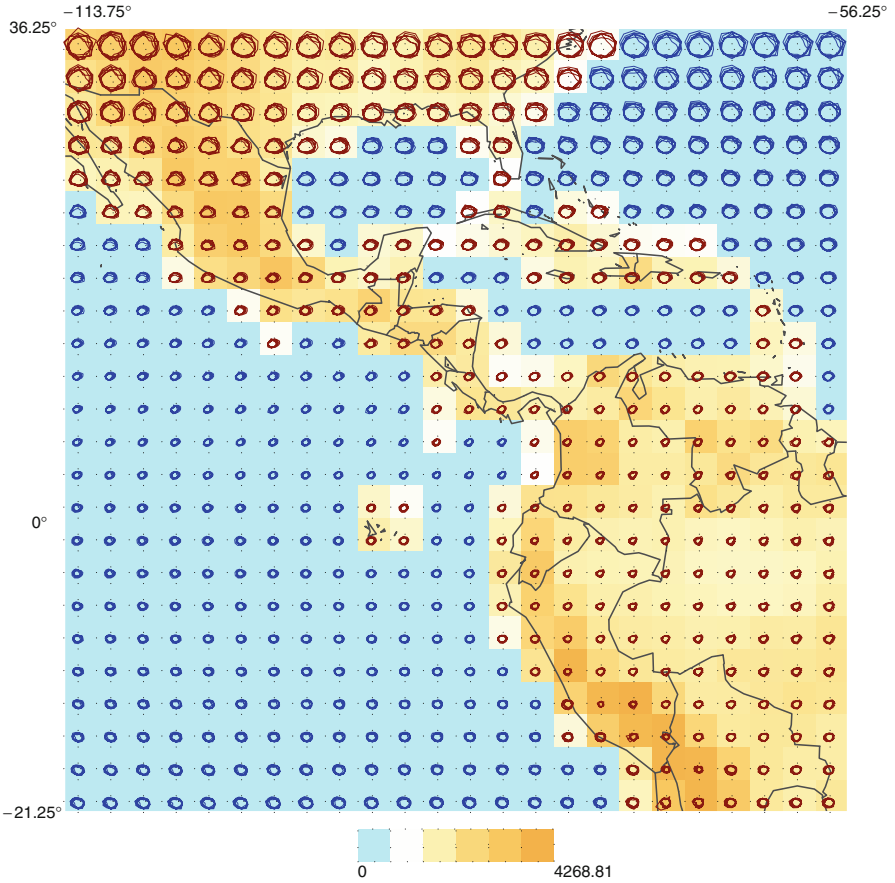

Fig. 5 Ozone abundance time series in polar coordinates plotted over a map of Central America. The background is the elevation of the region (in meters) with *lighter orange* denoting lower elevation, *darker orange* denoting higher elevation, and *blue* representing zero elevation. Each polar time series is built according to the convention of $2\pi = 1$year (see Fig. 4). The center of each geographic location is delineated by the *dotted lines*. The *line color* of the data points is used to further differentiate elevation with *blue* representing zero elevation and *dark red* representing greater than zero elevation

represents the month and the length of the cone is proportional to the average monthly frequency. Color was also used to further distinguish each month and is not related to the length of the cone. The colors were chosen to distinguish easily between the 12 months: green tones are used for spring, red for summer, purple for fall, and blue for winter months.

Returning to our Data Exposition data set, in Fig. 7 we used circular histograms to display the average monthly frequency of maximum ozone abundance at each geographic location in our grid of graphs. From this Figure, we see that maximum ozone abundance occurs in spring (March–May) in the northern part of Mexico and in the southern part of North America.[2] We also see that the month in which maximum

[2] Monks (2000) has shown this trend in the northern hemisphere in general.

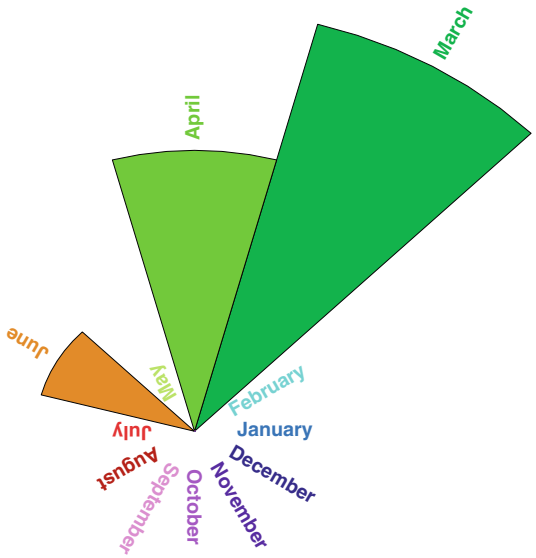

Fig. 6 Circular histogram for ozone abundance maximum. A circular histogram illustrates the frequency of a certain periodic event over several periods. The larger the frequency the longer the "cone". For example, in this figure, the average monthly frequency of maximum ozone abundance was 1/2 in March (observed in March in 3 years out of six), 1/3 in April (observed in April in 2 years out of six), 1/6 in June (observed in June only in 1 year out of six), and zero in other months (never observed in any other month of the year during the period of 6 years). Like polar coordinate plots, angle in circular histograms increases counter-clock wise. The colors were chosen to distinguish easily between the 12 months: *green tones* are used for spring, red for summer, purple for fall, and *blue* for winter months

ozone abundance occurs most frequently shifts from March/April to predominately July to September/October as latitude changes from the northern to the southern hemisphere. Lastly, the month of maximum ozone abundance is very consistent within some geographic regions, while quite variable in others.

3 Exploring the data from a local view: displaying the relationships between multiple variables

The more succinct methods of data display presented in Sect. 2.2 permit the investigator to decide which geographical areas and variables warrant further investigation. As seen in the previous section, the global perspective of ozone indicated several local areas with patterns of ozone seasonality that differed from the whole or surrounding patterns. Because no atmospheric variable operates in a vacuum, it is important to examine ozone's relationship with the other variables at these local areas in order to more fully understand this seasonality trend.

Traditional methods for displaying the relationships between multiple continuous variables involve creating individual scatter plots of all pairwise combinations between the variables of interest. A more concise display is a scatterplot matrix (Chambers et al. 1983), which uses a matrix of scatterplots to display all the pairwise relationships

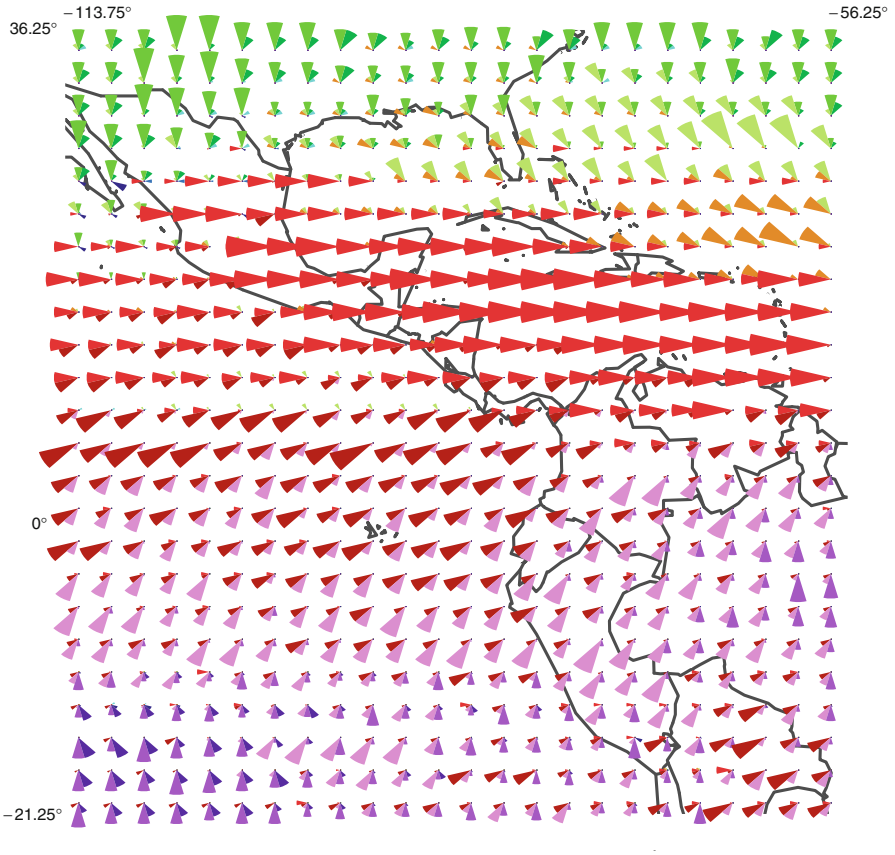

Fig. 7 Circular histograms of ozone abundance maximum over a map of Central America. See the caption of Fig. 6 for further details

between a defined set of variables. The individual scatterplots comprising the scatterplot matrix are referred to as panels which are labeled by their Y ˜ X (Row ˜ Column) variable names. Each pair of variables appears twice in a scatterplot matrix, once below the main diagonal and again as a mirror image above the main diagonal. In addition, all panels in the same row have identical y-scaling and all panels in the same column have identical x-scaling, thus making it easier to trace a point of interest from one panel across the other panels.

Based on the grids of graphs in Figs. 5 and 7, we chose a subset of geographic locations, elevations and time points to examine the relationships between ozone abundance, near-surface temperature, and low and high cloud coverage using several scatterplot matrices. We chose several locations at similar elevations but different latitudes/longitudes for analysis, and we then chose several points at different elevations (0, 2,000, 4,000 m) but somewhat similar latitudes or longitudes or both. These relationships were also examined for a fixed month (April) across all 6 years of available data and across all months for a fixed year (1996). Within each panel, different colored points were used to distinguish between the different geographic locations and

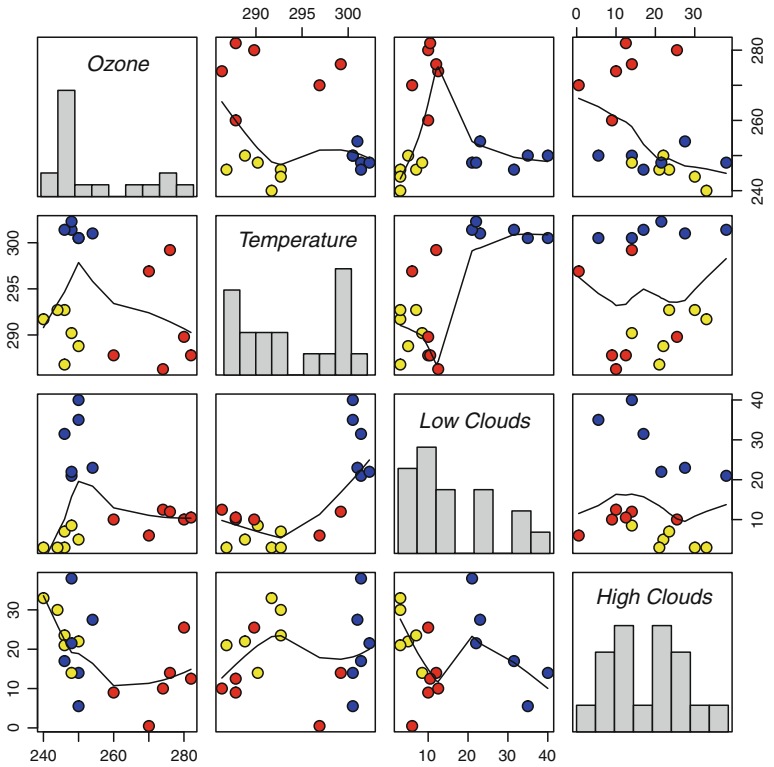

Fig. 8 Relationships between selected atmospheric variables for the month of April across all 6 years of data at an elevation of 2,000 m. The *blue circle* represents grid point A (24, 21) on the grid of graphs in Fig. 1 (counted from the top-left corner), the *yellow circle* represents point B (17,15), and the *red circle* represents point C (6,6)

elevations. A LOWESS (locally weighted scatterplot smoothing) curve was also added to each panel to display the relationship between x and y. Lastly, histograms were also plotted along the diagonal of the scatterplot matrix to illustrate the distribution of each variable.

For illustrative purposes, we decided to include only two of the generated scatterplot matrices. First, Fig. 8 plots the atmospheric data for the month of April across all 6 years of data for geographic points located at an approximately constant elevation of 2,000 m. The three geographic locations were grid points (24,21), (17,15), and (6,6) (counted from the top-left corner; corresponding to locations A, B, and C, respectively in Fig. 1). Secondly, Fig. 9 plots these same variables for April across all 6 years of data for geographical points located at approximately 0, 2,000, and 4,000 m, corresponding to grid points (4,22), (24,21), (21,18) (counted from the top-left corner; corresponding to locations D, A, and E, respectively in Fig. 1). From Fig. 8, we see that ozone abundance at two of the three grid points A (24,21) and B (17,15), i.e., the blue and yellow circles, is fairly constant across differing values of temperature and low and high clouds. In contrast, the ozone abundance at grid point C (6,6), the red circle, is quite erratic. In Fig. 8, we also see from panels above the main diagonal that

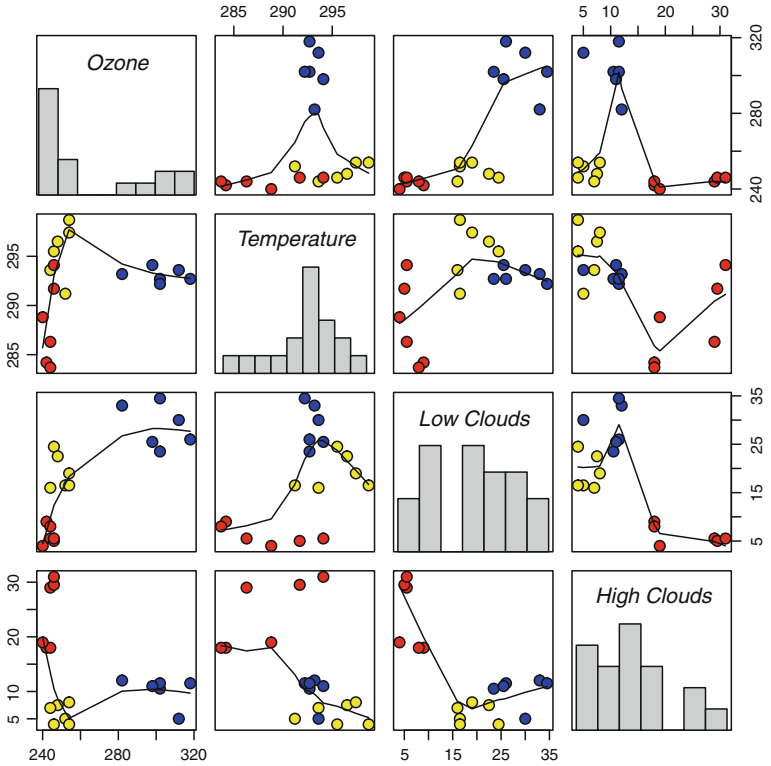

Fig. 9 Relationships between selected atmospheric variables for the month of April across all 6 years of data at elevations of 0, 2,000, and 4,000 m. The *blue* circle represents the grid point D (4,22) on the grid of graphs in Fig. 1 (counted from the top-left corner) at 0 m elevation, the *yellow circle* represents point A (24,21) at 2,000 m, and the *red circle* represents point E (21,18) at 4,000 m

no apparent relationships appear to exist between ozone abundance, temperature, or low and high clouds in the month of April at these three grid points. We see similar ambiguity in Fig. 9; ozone abundance at sea level, the blue circle, was the highest and fluctuated the most even with little change in temperature, or low or high clouds.

These illustrated scatterplot matrices are only two of a nearly infinite number that can be generated from this data set. The choice of geographic locations, elevations, and time points to plot depends heavily on the choice of the individual examining the data, as does the selection of the variables and the inclusion of additional graphical components (e.g., the plot smoothed estimates and histograms). Indeed, deciding on the group of scatterplot matrices that would best display the data set's most compelling or desired features would require much trial and error.

4 Summary/discussion

When presented with the Data Exposition challenge, we chose to focus on the process of determining the data's display, rather than making conclusions from the data and pointing out the data's features and/or flaws. Our goal was to use common graphical

methods in a new combination in order to incorporate as many components of the presented spatial, temporal, and multivariate data set as possible.

The methods we employed have certain advantages. The grid of graphs provided a global view of a large geographic area while also incorporating elevation. Polar coordinate plots clearly illustrated variability in seasonality, variability from year to year at a single geographic location, as well as across multiple geographic locations. Circular histograms allowed for one display of the frequency and seasonality of a temporal event of interest. Using the grid of polar coordinate plots or circular histograms to view the data globally also allowed one to more easily determine which geographic locations warranted further investigation. In turn, the scatterplot matrices provided an easy way to compare relationships between variables, incorporating multiple sources of variability (space, elevation, and time).

These methods also had their limitations. Although the grid of graphs can encompass any geographic area; the larger the area, the smaller each individual graph is when plotted on standard size paper, making it difficult to distinguish details. In addition, the polar coordinate plots may not be the optimum tool to capture the time trend of a variable because of the inability to distinguish between years. Additional graphical modifications to this method may be able to better accommodate a time trend.

The aforementioned advantages and disadvantages present us with additional challenges. Is there a potential to detect long-term changes in seasonal cycles using these methods? Would detection of long-term trends be better served by time series decomposition? And lastly, is it possible to use these methods with non-geographic and non-atmospheric data, such as electrocardiograms which are periodic?

Acknowledgments We would to thank Peggy Schuyler, R.N., for the editing assistance she provided.

Appendix: definition and example application of a circular histogram

To better understand the idea of a *circular histogram* we further develop the concept of time as a spiral, introduced in Sect. 2.2. Without loss of generality, let's assume that our time grid is even—the time interval between any two consecutive observations is the same and every cycle has the same number of observations. Let's also assume that one cycle is 1 year, which maps 1 year into the interval $[0, 2\pi)$. Every nth year is then mapped into the interval $[2\pi(n-1), 2\pi n)$. Given an even time grid, every time point j of an nth year is mapped into point $2\pi(n-1)+jk$, where k is the time interval between two consecutive observations. Thus, for monthly observations, $k = 2\pi/12$.

From this, we can define (Definition 1) a *time spiral*, $\Psi_{n,p}$, as a set points:

$$\begin{array}{ccccc} \Psi_{n,p} = \{\, 0, & \ldots, & jk, & \ldots, & (p-1)k, \\ 2\pi, & \ldots, & 2\pi + jk, & \ldots, & 2\pi + (p-1)k, \\ \ldots, & \ldots, & \ldots, & \ldots, & \ldots, \\ 2\pi i, & \ldots, & 2\pi i + jk, & \ldots, & 2\pi i + (p-1)k, \\ \ldots, & \ldots, & \ldots, & \ldots, & \ldots, \\ 2\pi(n-1), & \ldots, & 2\pi(n-1) + jk, & \ldots, & 2\pi(n-1) + (p-1)k\,\} \end{array} \tag{1}$$

where $p, n \in N$, $p \geq 2$, $n \geq 1$, $k = \frac{2\pi}{p}$.

We can also define additional concepts:

Definition 2 An ith *partial time spiral* of a *time spiral* $\Psi_{n,p}$ is

$$\Psi_i = \{2\pi i, 2\pi i + k, \ldots, 2\pi i + jk, \ldots, 2\pi i + k(p-1)\} \tag{2}$$

where $i \in N, 0 \leq i \leq n-1$.

Definition 3 Suppose we have a function $V(\psi)$ defined on a *time spiral* $\Psi_{n,p}$. A *partial circular histogram for* $V(\psi)$ *maximum on* Ψ_i (ith *partial time spiral*) is

$$h_i(\phi) = \begin{cases} \frac{1}{\text{N of maxima}} & \text{if } (2\pi(i-1)+\phi) \in \arg\max_{\psi \in \Psi_i}(V(\psi)) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\phi \in \Phi$.

Definition 4 A *time circle* is a set

$$\Phi = \{0, k, 2k, \ldots, jk, \ldots, (p-1)k\} \tag{4}$$

where $p \in N$, $p \geq 2$, $k = \frac{2\pi}{p}$.

Definition 5 A *circular histogram for* $V(\psi)$ *maximum on* $\Psi_{n,p}$ (a *time spiral*) is

$$H(\phi) = \frac{\sum_{i=1}^{n} h_i(2\pi(i-1)+\phi)}{n} \tag{5}$$

where $h_i(2\pi(i-1)+\phi)$ is a *partial circular histogram for* $V(\psi)$ *maximum on* Ψ_i *and* $\phi \in \Phi$.

Let's now re-visit the example introduced in Sect. 2.2 and illustrated in Fig. 6. Suppose we have 3 years of monthly ozone concentration observations from January 2000 to December 2002. Let's assume that during the year 2000 ozone reaches its maximum in March and June. During the year 2001 there is only one ozone peak in March. In year 2002 the ozone concentration peak is observed in April.

Our data has 3 years of monthly observations, therefore $n = 3$, $p = 12$, and $k = \frac{\pi}{6}$. Thus, the *time spiral* for our data is

$$\Psi_{3,12} = \left\{ \begin{array}{c} 0, \quad \frac{\pi}{6}, \quad \ldots, \quad \frac{j\pi}{6}, \quad \ldots, \quad \frac{11\pi}{6}, \\ 2\pi, \; 2\pi + \frac{\pi}{6}, \ldots, 2\pi + \frac{j\pi}{6}, \ldots, 2\pi + \frac{11\pi}{6}, \\ 4\pi, \; 4\pi + \frac{\pi}{6}, \ldots, 4\pi + \frac{j\pi}{6}, \ldots, \; 4\pi + \frac{11\pi}{6} \end{array} \right\} \tag{6}$$

The *partial time spiral* is

$$\Psi_i = \left\{2\pi i, 2\pi i + \frac{\pi}{6}, \ldots, 2\pi i + \frac{j\pi}{6}, \ldots, 2\pi i + \frac{11\pi}{6}\right\} \tag{7}$$

where $i \in \{0, 1, 2\}$.

The *time circle* is

$$\Phi = \left\{0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}, \frac{6\pi}{6}, \frac{7\pi}{6}, \frac{8\pi}{6}, \frac{9\pi}{6}, \frac{10\pi}{6}, \frac{11\pi}{6}\right\} \tag{8}$$

The *partial circular histograms* of maximum are

$$h_{2000}(\phi) = \begin{cases} \frac{1}{2} & \phi \in \left\{\frac{2\pi}{6}, \frac{5\pi}{6}\right\} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$h_{2001}(\phi) = \begin{cases} 1 & \phi = \frac{2\pi}{6} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$h_{2002}(\phi) = \begin{cases} 1 & \phi = \frac{3\pi}{6} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

And the resulting *circular histogram* of maximum is

$$H(\phi) = \begin{cases} \frac{1}{2} & \phi = \frac{2\pi}{6} \\ \frac{1}{3} & \phi = \frac{3\pi}{6} \\ \frac{1}{6} & \phi = \frac{5\pi}{6} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

References

Carslaw D (2005) On the changing seasonal cycles and trends of ozone at mace head, ireland. Atmos Chem Phys 5:3441–3450

Chambers J, Cleveland W, Kliner B, Tukey P (1983) Graphical methods for data analysis. Chapman and Hall, London

Lehman J, Swinton K, Bortnick S, Hamilton C, Baldridge E, Eder B, Cox B (2004) Spatio-temporal characterization of trophspheric ozone across the easter united states. Atmos Environ 38(26):4357–4369

Lindskog A, Beekmann M, Monks P, Roemer M, Schuepbach E, Solberg S (2003) Tropospheric ozone research—tor-2, international scientific secretariat-iss, gsf—national research center for environment and health, munich, germany. TOR-2 final report

Monks P (2000) A review of the observations and origins of the spring ozone maximum. Atmos Environ 34:3545–3561

R Development Core Team (2009) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, http://www.R-project.org, ISBN 3-900051-07-0

Scheel H, Areskoug H, Geib H, Gomiscek B, Granby K, Haszpra L, Clasinc L, Kley D, Laurila T, Lindskog A, Roemer M, Schmitt R, Simmonds P, Solberg S, Toupance G (1997) On the spatial distribution and seasonal variation of lower-troposphere ozone over europe. J Atmos Chem 28:11–28

Vukovich F (1994) Boundary layer ozone variations in the eastern united states and their association with meterological variations. J Geophys Res 99(D8):16,839–16,850

Wilkinson L (1999) The grammar of graphics. Springer, New York