36-315: Statistical Graphics and Visualization

Homework 7
Date: February 26, 2003 **Due: start of class March 3, 2003**

---

1. In homework 6, you showed that melanoma in Connecticut can be explained by an upward trend plus an oscillation plus noise. This time you will give a deep analysis of the oscillatory component compared to the sunspot data. As before, turn in all of your code, for all parts.

   (a) Download `hw6.csv` and source `lab7.r`. As in homework 6 and lab 7, decompose the melanoma series into a trend, oscillation, and noise. Make sure that you do not over-smooth the oscillatory component. One way to check this is to graph the noise with the same span—there should be minimal structure in the prediction line. Graph the oscillatory component with a connected line, in the correct aspect ratio.

   (b) Plot the auto-correlation function of the oscillatory component. What is oscillatory component's period, and how strong is the periodicity? Indicate where on the plot this is shown.

   (c) Make a spiral plot and calendar plot of the oscillatory component, using your estimated period. Make them easily-readable.

   (d) Find the cycle which is most out of sync with respect to the others. ("One of these things is not like the others...") Identify it on both plots and give a brief explanation of how both plots show this.

   (e) The melanoma oscillation is related to the sunspot number. Plot the cross-correlation function between the oscillatory component and the sunspot data. This will tell you the precise lag between the two. The cross-correlation is computed similar to `acf`, as follows:

   ```
   ccf(ts1,ts2)
   ```

   where `ts1` and `ts2` are vectors, e.g. columns of your frame. What is the lag between sunspots and melanoma? Indicate where on the plot this is shown.

2. In this problem, you will use the rate-of-change transformation to understand population growth in the U.S. A matrix of population data is in `hw7.csv`. Each column is a state and each row is a decade during which a census was taken. The functions you need are in `lab7.r`.

   (a) Start by making a line chart where the horizontal axis is time. This orders the states by total population, which obviously favors larger states.

   (b) Now change it into a growth chart. Plotting the growth since 1790 is more fair to small states, except Ohio which wasn't a state until 1803. Still, there isn't much to get from this plot except a basic ordering.

(c) Now make a new matrix containing the rate-of-change transformation of the population growth curves. This is done using the function `rate`, which takes a matrix as input and returns a new matrix where each number is the instantaneous percent change between rows.

(d) Make a line chart of the rate-of-change for each state over time. Use the option `ylim` to zoom in on the area between -10% and 50% growth rate, and choose a good aspect ratio. If you can't print in color, you can hand in the plot in black and white, though it obviously won't be as readable.

(e) Based on the similarity of their rate curves, pair up the states. For each of the three pairs, point out briefly what is the main similarity between the growth rates of the two states, and the main difference. Refer to dates but otherwise avoid using numbers in your description.

(f) After 1940, one state is not like the others. Which is it and how is it different? (Notice that this is much harder to see in the first two plots.)

3. The line chart entitled "The Aging American Electorate", at `http://lilt.ilstu.edu/gmklass/COW3/archive/Feb1902voting.htm` and reprinted on the next page, tries to depict how the age distribution of American voters has radically changed over time. Unfortunately, our perception of this trend is inhibited by certain choices made in the design of the plot. What are they and what could be done to emphasize the message?

4. Florence Nightingale, in her campaign for improvements to military hygiene, developed some novel plots of statistical data. One of these, the wedge plot, appears at `http://www.math.yorku.ca/SCS/Gallery/images/coxcomb.gif` and is reprinted on the last page. It depicts a time series by the size of wedges arranged in a circle. The circle connects the first time with the last time. More than one series can be shown by stacking the wedges. Critically evaluate this graphical method, using the concepts of data-ink, visual encoding, and visual connection discussed in class.