

36-315: Statistical Graphics and Visualization

Homework 6

Date: February 19, 2003

Due: start of class February 24, 2003

1. In this problem, you will take a seemingly complex time-series and show that it is actually pretty simple. The time series is the yearly incidence rate of melanoma in Connecticut. Download `hw6.csv` and source `lab6.r` just like in lab. The problem includes writing R code. Be sure to turn in all of your code, for all parts.
 - (a) Graph melanoma versus time using connected dots. Adjust the aspect ratio according to the 45° rule, which should be done for all remaining plots. The upward trend is pretty disturbing (the rate is normalized for population), but there is more to the data than that.
 - (b) Decompose the melanoma series into a linear trend and its residual. Start by graphing a trend line on top of a scatterplot. Use `span=1`.
 - (c) Now extend the frame with a column for the residual. Graph the residual series with a trend line of appropriate smoothness. The series has length 37, so choose a span of the form $k/37$. The result should be a smooth oscillation, without small bumps. You will probably need to adjust the aspect ratio for this plot.
 - (d) Now extend the frame to include the second fit and its residual. The first fit is the “trend component”, the second fit the “oscillatory component”, and the final residuals are the “residual component”. Graph the residual component using your favorite method. There is a pattern among the large residuals. What is it?
 - (e) Melanoma is mainly caused by solar radiation. The column `sunspot` is the number of sunspots observed on the sun during that year. Make a superposition of `sunspot` and your oscillatory component. These series have very different scales, so you will need to standardize them first. The function `match.quantiles` will take a matrix whose columns are time series and returns a new matrix where the columns are standardized by a scale and shift to match the first column.
 - (f) What connection between sunspots and melanoma is shown on the plot?