36-315: Statistical Graphics and Visualization

Lab 6

Date: February 18, 2003 **Due: end of lab**

Interspersed throughout this lab are some useful thought questions. You will be asked about them at check-off.

1. Download the files for this lab from the course web page.
2. Open a Word document to record your work.

Start R

3. `Start -> Programs -> Class software -> R 1.5.1`
4. Set the working directory to `My Documents`:

   ```
   File -> Change dir...
   ```

5. Load the special functions for this lab:

   ```
   source("lab6.r")
   ```

Load the data

6. `ozone = read.csv("lab6.csv")`

   `ozone` is a matrix with three variables:

day	Day of measurement (day 1 was May 1, 1974)
ozone.stamford	Ground-level ozone (ppb) in Stamford, Connecticut
ozone.yonkers	Ground-level ozone (ppb) in Yonkers, New York

Stamford ozone

7. Extract just the matrix pertaining to Stamford, as follows:

   ```
   frame = ozone[,1:2]
   frame = na.omit(frame)
   ```

   The last step was to remove days with missing values.

8. Using `frame`, make a simple scatterplot of ozone versus day. *What structure can you make out in this plot, specifically regarding trend, oscillation, and outliers?*
9. Make a new plot using vertical lines. This is done by giving `plot` the optional argument `type="h"`. This plot may suggest to you that the aspect ratio is too large (narrow). Try making the window short and wide. *What structure do you see now?*

10. Make a third plot using connected dots. This is done by giving `plot` the optional argument `type="o"`. *This and all further plots should be made with a good choice of aspect ratio. Does this plot help your perception of the structure?*

11. Use `predict.plot` to make a scatterplot of dots with a prediction line. Use a span of 10/153. This function can set the aspect ratio for you, by giving it the optional argument `asp="auto"`. (Is it similar to the aspect ratio you chose by hand?) *Does this plot help your perception of the structure?*

12. The prediction line made by `predict.plot` can be accessed using the function `smooth`:

    ```
    fit = smooth(ozone.stamford~day,frame,span=10/153)
    ```

    Use this to add new columns representing the prediction line and the residuals from it:

    ```
    frame = extend.with.fit(frame,fit)
    ```

    Type `frame[1,]` to see what columns were added.

13. Make a plot of the residual versus day, using vertical lines. *What outliers stand out in this plot? Are they the same as the outliers you identified in the previous plots?*

Stamford versus Yonkers

14. Now let `frame` be the full dataset:

    ```
    frame = ozone
    ```

15. For each place, fit a prediction line and add new columns to frame. (Just repeat the commands above for `ozone.stamford` and `ozone.yonkers`.)

16. Extract a matrix of the smoothed ozone levels, as follows:

    ```
    m = frame[,c("ozone.stamford.smooth","ozone.yonkers.smooth")]
    m = as.matrix(m)
    ```

17. Using the preceding matrix, make a line chart of the smoothed ozone levels, with time as the horizontal axis. *Is there a correlation between the ozone levels at Stamford and Yonkers?*

18. Make a growth chart of the smoothed ozone levels:

    ```
    growth.chart(t(m))
    ```

    This allows you to compare the height of the oscillations, as a percentage of the starting value. *Does ozone oscillate in the same way between the two cities?*

19. Show us your graphs.