STEVENS INSTITUTE of TECHNOLOGY
School of Business

FE 582

Lecture 3: Case Study: Data and Web Technologies

Dragos Bozdog

Exercise: Modeling Runners' Times

http://cherryblossom.org/

Exercise: Modeling Runners' Times

Topics:

- Scrape simple Web pages for text content
- Use regular expressions to extract and clean messy data from pre-formatted text tables and to create unique identifiers for matching records that belong to the same individual.
- Employ statistical techniques to identify bad data and to confirm these problems have been corrected.
- Visualize data that have a large number of observations (~150,000 records).
- Gain experience with the R formula language for plots and modeling.
- Fit piecewise linear models using least squares and non-parametric curves using local averaging.
- Compare data structures, e.g., a data frame and a list of data frames, for holding and working with longitudinal data. This includes the application of 'apply' functions such as tapply(), mapply(), sapply(), and lapply().

STEVENS
INSTITUTE of TECHNOLOGY
School of Business

Thank You

Dragos Bozdog

For academic use only.