Ian Nisbet

OCN 506

HW 3

Example Workflow:

In traffic engineering, we analyze a lot of data collected from traffic sensors in the road. One example of this is loop detector data collected from inductor wire loops embedded in the road surface. These devices can produce very useful data, such as vehicle speed, volume of traffic, and headway between cars. Data from this device can be streamed in real time, but is often collected and saved in a database. To analyze the data, it is often necessary to work through these common steps:

Goal: extract collected data into workable format, pre-process data, analyze, and output results in report ready format.

* Collect data (1-3 hours)
* Access the database where the data is stored (often Microsoft SQL server)(10 minutes).
* Export raw data to CSV or similar data type (10-30 minutes).
* Import data into R for processing and analysis (10 minutes).
* Pre-process data by removing errors or null values from the data frame. (30-60 min)
* Run R code to analyze the dataset for information we are looking for, such as linear regression or plotting data (60-90 minutes).
* Work out any code bugs for unexpected results (30-60 minutes).
* Export results and tables to include in report or research paper, often in LaTeX format (60-90 minutes).

It seems like some of this could be automated with shell scripts to speed up the process.