Virtual Data Intensive Summer School
July 8-10, 2013

Agenda

The Virtual Data Intensive Summer School will be held 8:00-2:00 PDT (11:00-5:00 EDT) on July 8 through July 10. To accommodate participants from different time zones, a short lunch break will be held at roughly 11:00-11:30 PDT (2:00-2:30 EDT) each day. All times below are given in Pacific Daylight Time and are subject to minor changes.

Monday, July 8

8:00-9:00	Robert Sinkovits, San Diego Supercomputer Center Introduction to summer school and basics of data intensive computing
9:00-10:00	Steve Tuecke, University of Chicago Globus Online for Research Data Management
10:00-10:15	Break
10:15-11:15	Yoav Freund, University of California, San Diego MapReduce, Hadoop and Spark
11:15-11:45	Lunch
11:45-2:00	Richard Marciano, University of North Carolina Jeff Heard, Renaissance Computing Institute (RENCI) Data management

Tuesday, July 9

8:00-11:00	Chris Fariss, University of California, San Diego Introduction to R and statistical analysis of data using R (break at approximately 9:30)
11:00-11:30	Lunch
11:30-12:00	Chris Fariss, University of California, San Diego Statistical analysis of data using R (continued)
12:00-2:00	Charles Elkan, University of California, San Diego The landscape of analytics: a personal view

Wednesday, July 10

8:00-11:00	Amy Szcepanski, University of Tennessee, Knoxville Introduction to Visualization with R (break at approximately 9:30)
11:00-11:30	Lunch
12:00-2:00	Dean Abbott, Abbott Analytics Inc. Text mining

Preparing for the virtual summer school

Several of the instructors have requested that you preinstall software on your laptop. Given the large number of participants and the compressed schedule, we ask that you comply and do this before the start of the summer school.

- R, version 3.0.1 (Good Sport) or later http://www.r-project.org/
- R packages: ggplot2 (version 0.9.3 or later), scales, ggmap, googleVis, and igraph
- Download uber cars data set from http://www.infochimps.com/datasets/uber-anonymized-gps-logs
- KNIME: http://www.knime.org/
- Google Chrome browser: https://www.google.com/intl/en/chrome/browser/

Prior knowledge of R is not required, but we do assume that you have some programming experience and familiarity with basic programming concepts (variables, arrays, loops, branching, etc.). You may find it helpful to acquaint yourself with basic R syntax ahead of time. Reading the first two chapters of the following online introduction is recommended http://cran.r-project.org/doc/manuals/R-intro.html

Prior knowledge of KNIME is not necessary, but you may find it helpful to familiarize yourself with the software. The KNIMEtech website contains very useful introductory material that will help you get started.

http://tech.knime.org/knime
http://tech.knime.org/installation-0
http://tech.knime.org/getting-started
http://tech.knime.org/screencasts
http://tech.knime.org/workbench