Empowering others to use data

Jeff Leek

@jtleek

www.jtleek.com

Data science training

- Internal talks
- Internal training sessions

https://www.datacamp.com/courses/big-data-revolution-r-enterprise-tutorial

DataCamp
We're hiring!

Courses Pricing Groups Create free account Sign in

Course

Big Data Analysis with Revolution R Enterprise

Revolution R Enterprise allows R users to process, visualize, and model terabyte-class data sets at a fraction of the time of legacy products without requiring expensive or specialized hardware. Introductory course for accomplished R users to experience the functionality of Revolution R Enterprise.

We like feedback; Let us know how you experienced this course!

Difficulty: Intermediate

Duration: About 8 hours

Participating: 13066

Start Course for Free

Or, take a subscription

Chapter 1: Introduction

Introduction to the RevoScaleR package that ships with Revolution R Enterprise. We discuss the challenges associated with dealing with Big Data and how the functions and algorithms in RevoScaleR address them.

Number of videos: **3** Number of exercises: **9**

Course given by:

Revolution Analytics

Revolution Analytics was founded in 2007 to foster the R Community, as well as support the growing needs of commercial users. Our name derives from combining the letter "R"

- Interactive documents
  - Data apps
  - Interactive visualizations

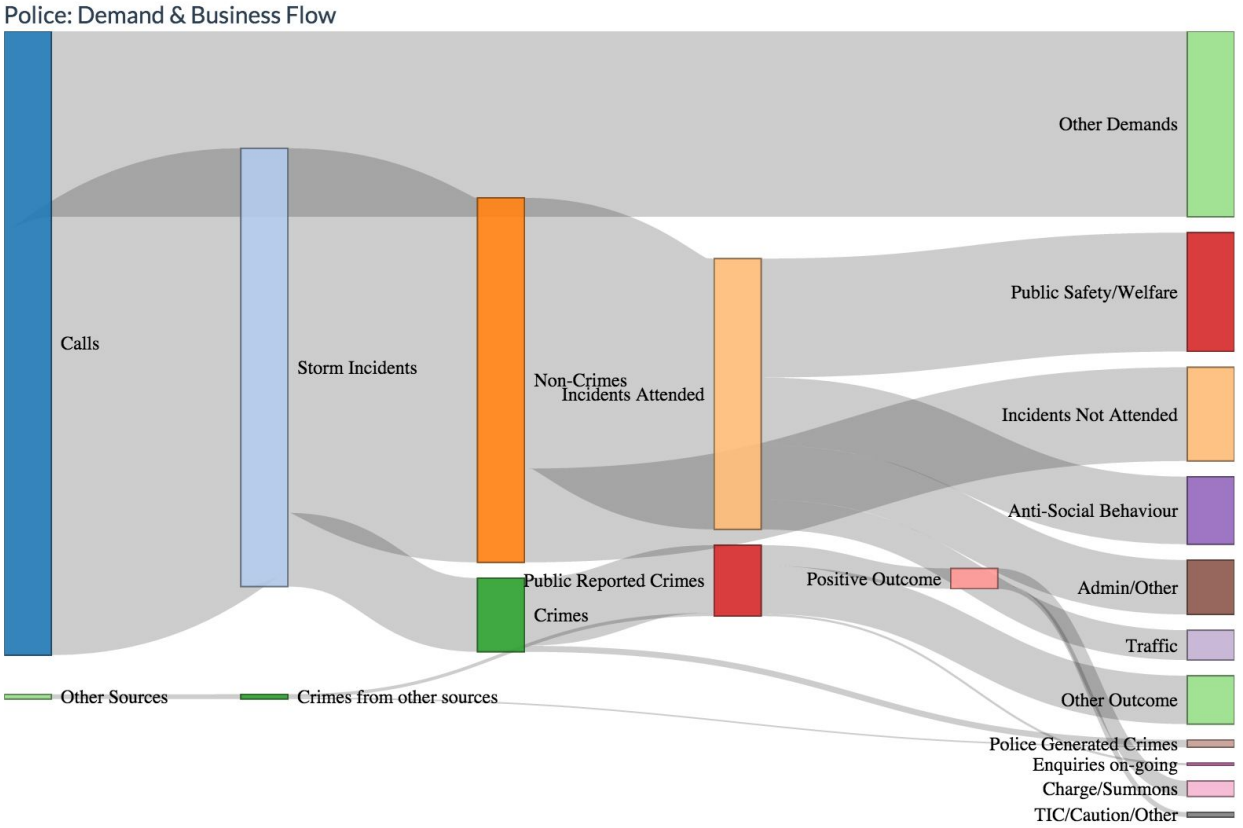
Police Supply & Demand Tool

This is a Pilot version of the Police Supply & Demand simulation tool.

Please note: the data contained in this demo version is *either artificial data or data already publicly available* (e.g. through http://data.police.uk/)

Methodology & Instructions

1. Analyse current crime patterns, detection & trends (including formal requests for specialist services):
    - Seasonality
    - Forecast
    - If needed for the current simulation run, adjust demand (forecast) to evaluate the impact of specific crime trends
2. Evaluate impact on resources
    - Historical analysis
    - Forecast-based
    - Adjusted forecast
3. Adjust Supply parameters to to fit Demand
    - Iterate
4. Analyse Outcomes



Pilot developed by Enzo Martoglio
for soprasteria http://www.soprasteria.com/

- Data idea evaluation
  - Data science is hard
  - Peer review from data team & you

http://data.washington.edu/incubator/

The Data Science Incubation Program

Summer 2015 Incubator Program: Data Science for Social Good

Previous Incubator Sessions

The eScience Institute is pleased to announce a new summer program: Data Science for Social Good (DSSG). The program is modeled after similar programs at the University of Chicago and Georgia Tech, but with elements taken from the eScience Institute's Data Science Incubator. Similar to the Incubator program, the goal of the DSSG is to enable new insight by bringing together data scientists and domain scientists to work on focused, collaborative projects. For the summer program, these projects will be designed to impact public policy for social benefit.

This year, the theme for DSSG projects will be Urban Science. Projects will involve analysis, visualization and/or software engineering of data from urban environments such as crime data government and NGO financials, housing data, cell phone GPS data, etc. Topic areas may include public health, sustainable urban planning, environmental protection, disaster response, crime prevention, education, transportation, governance, commerce, and social justice.

DSSG teams will be composed of:

- 4 students (see application link below) @ *5 days/week*
- 1 project lead, responsible for executing the research project (see proposal application below) @ *2 days/week*
- 1 public stakeholder, the organization or individual who provides the policy-relevant data and research question (see proposal application below) @ *1 hr/week*
- 1 eScience data scientist @ *2 days/week*

Our team of data scientists can provide expertise in state-of-the-art technology and methods in large-scale data manipulation and analytics (e.g., Hadoop, GraphLab, Myria, SciDB), cloud and cluster computing, statistics and machine learning, and visualization to help researchers extract knowledge from large, complex, and noisy datasets.

Any UW faculty, research staff, or non-UW scientist may submit a proposal to be considered for this program. DSSG projects are not "for-hire" software jobs -- each project will be led by representatives of the applicant's team working in collaboration with data scientists and the broader eScience community. An essential component of the application is the identification o