

## Basic Data Analytics Lab

This lab is a follow up to the Basic Data Analytics Techniques session. During that session there was a hands on demo of simple data analysis pipeline using the Chinook teaching database. This lab will give you the opportunity to apply the concepts and techniques discussed in that session to one or both of two new datasets.

### The goals of this lab are to:

- choose a dataset that is available for download from the internet (see below).
- contextualize the chosen dataset, gain an understanding of what the data is about and develop some plan for using the dataset to learn about the represented system.
- collect the data, make it available for analysis and prepare it for basic data analytics.
- carry out basic data analytics in R.
- carry out simple visualizations in R.

**Lab Structure:** You will be working in teams of 4 people. Each of you will have a role relating to a stage of the analysis process: conceptual analysis and problem definition, data loading, data preparation, data analytics, data visualization.

Carry out each stage of the data analysis process on your chosen dataset. The person with the relevant role at each stage will help to *lead* the process for that stage, but everyone will participate at each stage. **Note that everyone will collaborate for the data analytics part of the process.**

**Datasets:** You may choose to work with either or both of the following datasets.

The NorthWind teaching database:

<https://github.com/jpwhite3/northwind-SQLite3>

City of Ottawa open data on vehicle collisions (data for 2016 collisions):

<http://data.ottawa.ca/en/dataset/collisiondata2016>