



## WORKSHOP: Data Visualization and the Grammar of Graphics

### Presenter: Atul Sharma MD, FRCPC, MSc

Senior Biostatistician, Data Science Platform George & Fay Yee Centre for Healthcare Innovation  
318-1, CHOWN Building, 753 McDermot Avenue. Email: [atul.sharma@umanitoba.ca](mailto:atul.sharma@umanitoba.ca)

### Background:

In 2005, Leland Wilkison (Vice President Development, SPSS Inc.) published the *Grammar of Graphics* to describe the deep features that underlie all statistical graphics. In 2012, Hadley Wickham (Chief Scientist, RStudio) published *ggplot2: Elegant Graphics for Data Analysis*, a full implementation of Wilkinson's formal grammar for R (also available in SPSS, Stata, and SAS Proc IML Studio). With *ggplot2*, graphs are built from independent components that can be combined in many different ways. This makes *ggplot2* very powerful because the analyst is not limited to a set of pre-specified graphics, but can create new graphics that are precisely tailored to their specific problem. Moreover, because there is a simple set of core principles and very few special cases, *ggplot2* is easy to learn. The system is designed to work iteratively: You start with a layer showing the raw data then add layers of annotations and statistical summaries. This produces graphics using the same structured thinking used to design an analysis, which reduces the distance between a plot in your head and one on the page. Easy integration with R Markdown/R Studio also makes it easy to incorporate beautiful graphics into professional knitr reports.

### Structure:

Divided into two sessions, the first will provide a 2h introduction to the theory and application of Wilkinson's Grammar as implemented by Wickham, concentrating on real-world examples. The second session will be a 3h computer lab, to provide a hands-on experience with *ggplot2* and other graphical packages built on the formal grammar e.g. *ggmap*, which links *ggplot* to on-line map databases, like Google Maps and the OpenStreetMap project.

### Prerequisites:

While the lecture session may be of general interest, computer lab participants are expected to be familiar with R, particularly the base R plot functions and data frames. It is recommended that you bring a laptop with R and R Studio. For assistance, see *A (very) short introduction to R* by Torfs and Brauer, at <https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>.

A number of R packages will be used during the computer lab, which can be installed from the command line in a single step via:

```
install.packages(c("ggplot2", "plyr", "gridExtra", "mgcv", "RColorBrewer", "ggmap"))
```

### Workshop Leader:

Dr. Atul Sharma is a pediatric nephrologist and biostatistician, who recently joined the Biostatistical Consulting Unit as a senior consultant. As such, he brings a unique combination of clinical and statistical insights to methods of interest to clinical investigators.