Introduction

- The package "`ggplot2`" was developed by Hadley Wickham.
- It is based on the book "***Grammar of Graphics***" by Leland Wilkinson (SPSS, previously Systax).
- Simple set of core principles of how to render graphics, that can be built up iteratively, and edited later.
- `ggplot2` works in a layered fashioned, starting with the raw data at the base layer, and building up additional layers.

What is the grammar of graphics

- Wilkinson addresses the question of "what is a statistical graphic?"
- Mapping from data to aesthetic attributes (e.g. colour, shape and size) of geometric objects (e.g. points, lines and bars)
- Remark: geometric objects are referred to as "***geoms***".
- ***Stats:*** Statistical Transformations:
-
- ***Faceting:*** process where similar plots are generated for subsets of the data to allow comparison between them.

- Wickham's layered graphics approach is based on Wilkinson's grammar, but adapts it to fit it into the *R* environment.
- Ggplot2 is for static graphics only. For interactive graphics, try out `rggobi.`

Preliminaries

- Installation is straightforward.
- Just remember to have recent version of R.

```
>install.packages("ggplot2")
>library(ggplot2)
```

Book

- ggplot2 : Elegant Graphics for Data Analysis
- Author: ***Hadley Wickham.***
- Springer-Verlag, New York, 2009.

Websites

- **http://www.had.co.nz**
- **http://groups.google.com/groups/ggplot2**

Layered grammar

Layered grammar defines a plot as the combination of

- A default data set and a setting of mappings from variables to aesthetics
- One or more layers
- One scale for each aesthetic mapping
- A coordinate system
- The faceting specification

Layers

- Layers are responsible for creating the objects on the plot.
- Layers are composed of 4 parts
    1) Data / aesthetic mapping
    2) A statistical transformation (stat)
    3) A geometric object (geom)
    4) Position adjustment

Data sets in ggplot2

- The following data sets are embedded in the ggplot2 package. We will mostly use "diamonds" and "mpg" for this workshop .

diamonds:	Prices of 50,000 round cut diamonds
economics:	US economic time series.
midwest:	Midwest demographics.
movies:	Movie information and user ratings from IMDB.com.
mpg:	Fuel economy data from 1999 and 2008 for 38 popular models of car
msleep:	An updated and expanded version of the mammals sleep dataset.
presidential:	Terms of 10 presidents from Eisenhower to Bush W.
seals:	Vector field of seal movements.