Session 3 - Visualization

Jongbin Jung

January 9-10, 2016

Dependencies

- Latest version ($\geq$ 3.1.2) of R
  (*free* from https://www.r-project.org/)
- Latest version of Rstudio (also *free* from
  https://www.rstudio.com/)
- A bunch of *free* packages

```
install.packages('ggplot2')
```

Visualization: Introduction

- There is more than one framework for thinking about data visualization, e.g.,
  1. Mapping of vectors to 2D/3D surfaces
  2. Function of **inputs** that are given as varaibles of a data set, **geometries** and **aesthetics** that describe visual markings, and a **coordinate** system that defines the location of each marking
- The first approach is widely used in scientific visualization (e.g., `MATLAB`, classical plotting function in `R`), but doesn't scale well with data
- It helps your sanity to always think of data frames as a table where

```
Each column represents a variable/feature
Each row represents an observation/instance
```

Reference

- A great "cheat sheet" for data visualization with `ggplot2` is available for free at `https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf`