# Welcome to STA303/STA1002



clipartpanda.com

STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

# About the Class

- Review of p-values, hypothesis testing, confidence intervals, and all that
  - Using all of those concepts correctly
- Focus: analyzing data that doesn't conform to the assumptions of the Linear Model
  - Disentangling the effects of various covariates from each other on the variable of interest
  - Determining whether the data displays any pattern at all
  - Predicting future data
- Use of modern statistical tools for performing analysis
  - Computer simulation in R for analyzing various statistical techniques (you will use simulation even more if you work in statistics in the future)
  - Data Visualization in R
  - R Markdown for generating reproducible analysis

# About the Class

- Common requests
  - Working with financial data – coming up in Project 2
  - Data Science-flavoured assignment – also coming up in Project 2
  - Public Health, Psychology data, sports – still working on finding something suitable

# About the class

- **Tentative** schedule (to be finalized after I see how the first two lectures go, by Thursday)
  - Project 1: 10%, due Jul. 11
  - Project 2: 15%, due Aug. 1
  - Midterm: 25% on Monday Jul. 18
  - Exam: 50%, during the examination period

# About me

- Michael Guerzhoy is pronounced "Mike-ul ger-JOY"
- Office hours: R6-7, F3-4, or by appointment in BA5244

# About me

- Teaching full-time: mostly in Computer Science, but looking forward to teaching Stats!
- Data Science Consultant on the side
  - Clients include: CBC, Homestars.com, and people you've never heard of
- MSc in Stats (applied stats and statistical machine learning) and MSc in Comp Sci (machine learning/computer vision)
- Research interests: log-linear (and non-log-linear) models, hierarchical Bayesian models
  - More on that if time permits

# About you: Future Plans



Future Plans (incl. multiple responses)