# INTA 4803/8803 – Data Analysis & Visualization

Fall 2017
Meeting Time: Tuesday | 3:00-5:45
Meeting Location: Habersham 136
Instructor: Wes Stayton
Email: thwg@gatech.edu
Office: Habersham G16

## Class Objectives:

1. Practical Proficiency with R Language
2. Applied Knowledge of Basic Statistical Techniques
3. How to Analyze a Problem/Question → Creating a Statistical Model
4. Crafting an Intuitive Narrative
5. Visual Storytelling

This course is designed to expose you to statistical computing in a social science context. We will not focus on mathematical foundations or proofs. Rather, the course is designed to give an intuitive introduction to statistical modelling and, more importantly, teach you how to operationalize these models using the R programming language. Each stage of the analysis process is covered – data collection, data manipulation, model selection and execution, visualization of results, and reporting. You should leave the course feeling confident in your ability to perform quantitative analysis for both academic and "real world" problems.

The bulk of the course revolves around your term project/paper. We will begin working on the project early in the semester and check in multiple times before the final product is due. You will be expected to talk about your project *at least* three times throughout the semester and receive feedback from your colleagues. While lecture based instruction is important – there is no substitute for learning by doing.

## Deliverables

Term Paper/Project
- Proposal – Week 5……..10%
- 2x Feedback – Weeks 6 – 12……..10%
- Final Paper – Final Period……...30%
- Presentations – Weeks 16 & 17…..……...……..20%

Quizzes – Weeks 5 & 9……..……10%
Class Participation and Collaboration – Every Week…….20%

## Term Paper/Project

The term project will be a quantitative analysis of some problem or question that can be either academic or applied in nature. The goal of the paper is to get an idea of how to scientifically structure an inquiry and carry out the analysis. The statistics and programming are strait forward, the difficult part is thinking critically about how to approach the problem and use the tools you've acquired in the course to give a satisfying answer. Your paper should follow the classic format – abstract, literature review/background, model, data, methodology, results, and conclusion. To compliment your academic paper, you will give a presentation that is focused on conveying insight through visualizations. Think of the paper as being a tool for communicating with the academic community and the presentation as a tool for communicating with the practitioner community.

Your will turn in a proposal of your paper topic in week 4 and receive feedback the next week so you can get an early start. You will also give a quick 5 to 10 minute update on your project twice throughout the semester so you can receive further feedback from me and your class mates. The final paper is due electronically on the day of the final. You will give a final presentation, with Q&A, to the class in the final week. Your presentation should showcase your visualization skills.

## Materials

The assigned book for the course is Naked Statistics by Charles Wheelan, available here. Reference material will be posted on Tsquare. One of the most important skills for a data analyst is learning how to find information. If you have a question about R programming or statistics it will be useful to learn how to search Stack Overflow and R Bloggers.

# Schedule

Week 1 | August 22 – Class Introduction. What is Data Analysis?

Week 2 | September 29 – Intro to R

Week 3 | September 4 –Data Manipulation & Exploratory Data Analysis

Week 4 | September 11 – Visualization

Week 5 | September 18 **– Quiz 1 R Programming**

Week 6 | September 25 – Stats and Probability Review

Week 7 | October 2 – Regression

Week 8 | October 9 – Time Series and Panel Data

Week 9 | October 16 – **Quiz 2 Statistics**

Week 10 | October 23 – Text Analytics / Social Network Analysis

Week 12 | October 31 – Machine Learning

Week 13 | November 7 – No Class

Week 14 | November 14 – Big Data

Week 15 | November 21– No Class – Thanksgiving

Week 16 | November 28 – Presentations

Week 17 | December 5 – Presentations