# CS 591 – Fall 2015, Assignment 2
## Due by 10PM on Wednesday, September 16

Your second assignment is to present an analysis from an A/B test of two alternative website designs.

**Problem Statement:** Your Web development staff has conducted an A/B test of the conversion funnel on two versions of your website. For each user, they logged information that you requested, including whether or not the user converted, and the purchase amount, in the event that they did convert.

Your objective should be a justification to either roll out version B of your website, continue the experiment (specifying an estimate of how much more time will be needed), or stop the experiment and stick with website A. To present your results, first start with a basic interpretation of the data, and then move to a statistical analysis of the data. As appropriate, you may use any of the methods we've seen so far, including but not limited to, z-scores, T-tests, or Chernoff bounds. You may also want to take a look at the Wald and Wilson methods for computing confidence intervals. Keep your presentation short and to the point.

**Datasets:** The datafiles for websites A and B are linked from our class homepage, at WebsiteA.csv and WebsiteB.csv, respectively.

Each of the datafiles is a CSV (set of comma-separated entries, one per line), where each row corresponds to a customer visit. Website A has 10000 rows, and Website B has 8000.

Each row is a tuple of the format (timestamp, dwelltime, conversion, revenue) where timestamp is an arrival time in seconds, dwelltime is the time the user spent on site in seconds, conversion is a boolean indicating whether or not the user converted, and revenue is a revenue amount (in cents) for converting users. I kept everything integral for simplicity, but your routines may use floating point, of course.

**Sage:** An especially convenient way to do this type of analysis is to use Sage, alternatively, you can use your own preferred language. There is a bit of a learning curve to using Sage, but to get started, try the Sage tutorial at www.sagemath.org, and store your analyses on the SageMathCloud at cloud.sagemath.com: a nice way to run data analytics with Sage, Python, or R, and quickly share your results. I'll give a quick tutorial in class on Tuesday.

**Submission:** Submit your analysis (3-slide presentation or Sage worksheet) to me by no later than 10PM on Wednesday. We will take a look at a few of the visualizations to start class on Thursday.