# HOMEWORK 01 — MATH 295: DATA SCIENCE II

**Assigned: 2014-09-04**                                                **James P. Bagrow**
**Due:      2014-09-11**

## Instructions

Please read these instructions carefully before you begin!

Included with this file is a directory called `work/`. This is where you will include any code you have written to answer the questions. Inside `work` is a directory called `data/`. Your code should sit in the `work` directory and read those files from the `work/data/` subdirectory.

- **Show your work**. This means (i) provide the code you have produced by placing it in `work/`, (ii) next to any answers in your writeup include a parenthetical statement pointing out where in your code the answer was computed. For example: "(See plot_timeseries.py, lines 100-110.)"

**Your writeup**

Please make a file called `Homework01_<yourname>.pdf` containing your written answers to the below problems. Include the assignment number, date, your name, and label each problem clearly. Any figures you provide should also be in this file. You may use any tool you wish to make this file but you must convert the final output to a PDF. If you use Microsoft Word **please do not turn in a Word file**; export the file to PDF.

**To submit**

With your final `.pdf` writeup in `HW01/` and your code in `HW01/work/`, rename `HW01` to `HW01_<yourname>`, compress the directory and send me an email with `HW01_<yourname>.zip` attached.

## Problem 1

Included with the assignment are several text files. These files track the historical population of US states by year for 1900-1990.

### 1A

Write a script to process these data and load them into a data structure you can work with. What problems did you have to deal with when working with these files? (Ie, please provide a short written log or description of the work you needed to do here.)

### 1B

Plot the populations of Alaska and your home state over time. Plot the population of New England and the Southwest over time.

- Label your axes!

### 1C

What state showed the **greatest change** in population? Note that there is more than one way to quantify this. Please come up with at least **two** (meaningful) ways.

**Bonus**

Make a map of the US showing the change in population. Extra points for extra clever solutions!

# Problem 2

The file POLIO_Cases_1928-1969.csv contains the number of cases of polio over time.

### 2A

Write a script to load these data into a data structure you can work with. Please describe any issues you ran into when working with the file. In addition, what **sources of uncertainty** (or measurement error) do you think may affect these numbers?

### 2B

Using these data, provide some **descriptive statistics** and **graphical summaries** of the history of polio in the United States. Provide a brief description of this history, *with citations*, and describe how to see these trends in your statistics and graphics.

### 2C

Were some areas of the country harder hit by polio than others? Please demonstrate this **graphically** and do not forget to account for the total population (Problem 1).

<p style="text-align:center; font-size:1.5em;">Good luck!</p>