Data Journalism Bootcamp

Fall 2015

Course Description
We will learn how to investigate data-driven stories that catalyze change without knowing how to code. This hands-on lab course will start with the basics by honing students’ BS detector before covering the building blocks of data journalism through presenting findings for publication — be it in text, graphic or other form. Students will learn the basic math and spreadsheet skills needed to verify data and spot outliers that make a great story. Extracting and cleaning data and how to format findings for publication will also be covered. Students will leave knowing how to make spreadsheets do the dirty work of great journalism. Throughout the course, students will learn from published stories that changed lives and law and see how those journalists got their numbers.

Learning Outcomes and Objectives
This is a hands-on course. You will learn by doing. Each class we will cover a topic, require students to work in groups or individually using that skill and then complete an assignment building on that lesson for the following week. You should leave this course with a critical understanding of how data is used in journalism today and how to use those skills in your own journalistic career. Because the majority of lessons will be done in class, attending class is of utmost importance. Only medical absences are excused. No extensions will be given on homework unless approved by the professors prior to the due date.

Required Materials
IRE Student Membership: $25 Access to a spreadsheet program (Excel, Google Docs, Libre Office or Calc)

Grading
Grading is based on an A-F scale made up of these three parts:

50% — Attendance, participation in discussions and work, completing in class assignments
30% — Weekly assignments
20% — Final project

Final Projects
Each student will choose a raw data set (one they find and are interested in analyzing or one provided by the professors). Students will then “clean” and analyze the data in order to tell a story. The “story” can be in whatever format the student chooses be it print piece, static or interactive graphic or some other format.
Grades for the final project will be based on three, equal parts.

**Data Diary**: Detailed report of steps used in procuring, cleaning and analyzing the data for the project.

**Analysis**: Identified correct techniques and did those correctly. Techniques used, ed, e

**Presentation**: The final story, chart, etc.

---

**Course Schedule**

This is an outline. Everything is subject to change and assignments likely will be altered throughout the semester. Each class will begin with a tutorial, followed by hands-on work. Readings should be completed before that class period. Assignments will be handed out at the end of class and are due before next week’s class.

**Week 1 (Sept. 2): Introduction to core concepts -- Setting up computers**

Topic: Data provenance, source and how to make public records requests. We’ll figure out everyone’s familiarity with spreadsheets and basic math. Assignment: Objectives/Skills Survey, File an Open Records Request.

**Week 2 (Sept. 9): Basic Math**

**Readings to be completed by class**: The New Precision Journalism Chapter 2 (Figuring A Rate, Percent Difference, Times more v. Times As), WSJ Medicare Payments, NYT Medicare Payments

**Topic**: Mean, median and mode. Rates and percent differences. Basic excel skills through Medicare data. Assignment: Calculate the incident, fatality and fatal incident rates for all airlines listed, 8599 and 20002014

**Week 3 (Sept. 16): Advanced stats, functions**

Readings: Hospice Inc., How Software Increases Cab Revenue, Sarah Cohen’s Numbers in the Newsroom

**Topic**: Date conversions and comparisons, string functions, distributions through Hospice data. Also, understanding how to write about data in news stories. Assignment: None

**NO CLASS (Sept. 23): Yom Kippur**

**Week 4 (Sept. 30): Census, Charting, Advanced stats, functions continued**

Readings: Distrust Your Data, Using Census Microdata, Where We Came From (And Where We Went)

**Topic**: Margin of Error, advanced string functions using Census Micro Data (story), Basic Charting, FactFinder

**Assignment**: Pick a fact finder table and chart it.

**Week 5 (Oct. 7): Multiple calculations in a few easy steps**
Readings: PivotTables in Excel, At Chipotle, How Many Calories Do People Really Eat?, Hundreds of Police Killings are UnCounted in Federal Statistics, Topic: PivotTables and summarizing large data
Assignment: New York City Schools Data

Week 6 (Oct. 14): What went wrong? Understanding dirty data and how to clean it
Readings: Danger!! Look out for dirty data, The datadriven story part 3: bulletproofing and presentation, National Donors Pick Winners in State Elections, Topic: Understanding how and why data is wrong. More String Parsing! Bulk editing columns. Clustering to fix names. Assignment: Identify the top 10 states where the minimum number of opiate addicts exceed patient capacity

Week 7 (Oct. 21): Importing, Review of all functions and RegEx
Readings: Finding Data Online, What the Fox Knows, Topic: Importing data in different format spreadsheet, validate data is complete, summarize data using skills learned through the semester Assignment: Quiz

Week 8 (Oct. 28): Basic Scraping, Scraping PDFs Using Tabula, Comet Docs
Readings: Meet Your Web Inspector, Med Schools Flunk at Keeping Faculty Off Pharma Speaking Circuit, About the Dollars for Docs Data
Topic: Inspecting a web page, Google Scraping, DOM hierarchy, selecting DOM Assignment: Compare MUNI schedule adherence rates against published reports

Week 9 (Nov. 4): Basic Mapping
Readings: When A Map Isn’t A Map, Watch Darla Carmeron’s Lighting Talk: The end of Maps in 7 charts Topic: DIY map tools, loading data, geolocation Assignment: Use SFMTA adherence data from prior class to map routes by ontime rates

Week 10 (Nov. 11): Final Project/Guest Speakers
Readings: TBD Topic: Discussion of the final projects and QnA with professional data journalists. Assignment: Work on Final Projects

Week 11 (Nov. 18): Final Projects Workshop
Topic: Lab time to work on final projects and ask questions of Shane or Coulter regarding projects or any of the lessons from the semester.

Week 12 (Tue Nov 24): Present Final Project & Semester Wrap Up Topic: Student discuss their
final project.