Class Name: Introduction to Data Visualization

Course Number: J221

Semester/Year: Fall 2015

Location/Room: 106/Upper NG

Day/Time: Thursdays from 9.30am - 12.30pm

Instructor Name: Peter Aldhous

Contact Info: peter@peteraldhous.com
415 503 7323

Office Hours: Thursdays from 1.30pm - 5.00pm

Course Description:

This is a course in finding and telling visual stories from data. We will cover fundamental principles of data analysis and visual presentation, chart types and when to use them, and how to acquire, process and "interview" data. We will make interactive and static charts and maps using free software. There will be some coding, but no prior experience is required. The emphasis is on gaining practical skills that students can apply in a newsroom setting.

Introduction to Data Visualization, Fall 2015

This is a course in finding and telling visual stories from data. We will cover fundamental principles of data analysis and visual presentation, chart types and when to use them, and how to acquire, process and "interview" data. We will make interactive and static charts and maps using free software. There will be some coding, but no prior experience is required. The emphasis is on gaining practical skills that students can apply in a newsroom setting.

Weekly schedule

We will meet in **106/Upper NG** on Thursdays from 9.30am - 12.30pm. Your instructor, **Peter Aldhous**, will maintain office hours in **B1** from 1.30pm - 5.00pm, following each class. You are encouraged to arrange appointments to discuss your work.

Class time will also be scheduled for each of you to critique and lead class discussion of a recently published news graphic/interactive.

- ### Aug. 27: What is data?

  Categorical and continuous variables; basic operations for interviewing a dataset; sampling and margins of error; plotting and summarizing distributions; choosing bins for your data; basic newsroom math; correlation and its pitfalls; exploring differences between groups; scatter plots and box plots.

  - **Assignment:** make plots using ggplot2 web app; subscribe to data viz blogs etc to follow latest developments in visualization.

- ### Sept. 3: Data visualization: basic principles

  Encoding data using visual cues; choosing chart types to show comparisons, composition (parts of the whole) and connections; using color effectively; using chart furniture, minimizing chart junk and highlighting the story; avoiding pitfalls; good practice, including for interactive graphics.

  - **Assignment:** quiz on good practice in visualization and data analysis.

- ### Sept. 10: Interviewing data: exploratory graphical analysis

  We will use Tableau Public to explore and visualize World Bank data on neonatal deaths across the globe, creating an interactive online dashboard.

  - **Assignment:** make another dashboard from the same data, and combine into a Tableau story.

- **Sept. 17: Acquiring, cleaning and formatting data**

  Data search and download tricks, including Table2Clipboard and DownThemAll! Firefox plugins; manipulating urls and using APIs to acquire data; scraping data from the web with Kimono; cleaning data with Open Refine; converting data between different formats using Mr. Data Converter.

  - **Assignment:** data cleaning task with Open Refine.

- **Sept. 24: Interviewing data: using databases**

  Introduction to databases and Structured Query Language for manipulating data. We will use SQLite and the SQLite Manager Firefox plugin to explore data including drug company payments to doctors.

  - **Assignment:** write queries to return specified data.

- **Oct. 1: Let's apply what we've learned so far**

  We will obtain and process World Bank data on life expectancy and GDP per capita for the world's nations from the web, and then create a Tableau visualization inspired by this Gapminder video.

  - **Assignment:** download and process World Bank data on nations' carbon dioxide emissions per capita, and make a simple Tableau dashboard.

- **Oct. 8: Principles of mapping; introduction to final projects**

  Basic mapping principles: projections, geocoding, geodata formats; approaches to putting data onto maps, including choropleth maps, scaled symbols, hexagonal binning and cartograms. We will also discuss datasets made available for your final projects; you are encouraged, however, to suggest your own.

  - **Assignment:** submit a pitch detailing plans for your final project. You should also each arrange an individual appointments with your instructor over the coming week to discuss your plans.

- **Oct. 15: Making static maps and processing geodata with GIS software**

  We will use QGIS to make a multi-layered map. We will also learn how to use QGIS and its plugins to process geodata.

  - **Assignment:** make a simple map requiring a data join.

- **Oct. 22: Making interactive maps; processing geodata with SQL**

  We will use CartoDB and Leaflet to create interactive online maps. We will also use SQL and PostGIS, which is built into CartoDB, to process and analyze geodata.

  - **Assignment:** deliver finished online map; continue work on final project and deliver a progress report.

- ## Oct 29: No Class

  Instead, one-on-one meetings will be arranged with instructor to discuss progress with your final projects.

  - **Assignment:** refine plan for your final project.

- ## Nov. 5: Manipulating data and making graphics with R

  Introduction to R, R Studio and R packages including ggplot2 for visualization and dplyr for data manipulation.

  - **Assignment:** make specified graphics with ggplot2.

- ## Nov. 12: Coding interactive graphics

  We will use D3 to code from scratch a number of common chart types. This will be a challenging exercise, intended as an introduction to the huge possibilities offered by a JavaScript code library that powers many of today's most impressive online news visualizations.

  - **Assignment:** continue work on your final project.

- ## Nov. 19: Now let's make this easier: from R to interactive charts and maps

  We will explore R packages, including rCharts, that allow you to create JavaScript visualizations straight from your data with a minimum of coding.

  - **Assignment:** Continue work on your final project. Each of you should also an individual meetings with your instructor to refine your final projects, which are due on the last day of class.

- ## Nov 26: Thanksgiving — No Class

  - **Assignment:** continue work on your final project.

- ## Dec. 3: Other tools for online interactives

  We will review other tools for making online interactives.

  - **Assignment:** complete your final project.

- ## Dec. 10: Let's apply what we've learned; wrapping up

  As in week 6, we will obtain and process World Bank data, and make some interactive charts, but this time using code. We will conclude with a discussion of lessons learned, and next steps to continue developing your data manipulation and visualization skills.

Recommended reading

Alberto Cairo: *The Functional Art: An Introduction to Information Graphics and Visualization*
Nathan Yau: *Data Points: Visualization That Means Something*

Further reading/viewing will be recommended to support weekly class material.

Attendance

Unexcused absence from two classes will drop you one letter grade; a third unexcused absence will result in an F. Excused absences will be permitted only in extraordinary circumstances. Regardless of the reason for an absence, students will be responsible for any assignments due and for learning material covered in class.

Grading

Class participation, weekly assignments: **45%**
Final project: **45%**
Attendance: **10%**

Good manners

Students must turn off the ringers on their cell phones before class begins. Students may not check e-mail, social media sites or other websites during lecture portions of class or while working on class exercises.

Academic dishonesty and plagiarism

The high academic standard at the University of California, Berkeley, is reflected in each degree that is awarded. As a result, it is up to every student to maintain this high standard by ensuring that all academic work reflects his/her own ideas or properly attributes the ideas to the original sources.

These are some basic expectations of students with regards to academic integrity:

- Any work submitted should be your own individual thoughts, and should not have been submitted for credit in another course unless you have prior written permission to re-use it in this course from this instructor.
- All assignments must use “proper attribution,” meaning that you have identified the original source of words or ideas that you reproduce or use in your assignment. This includes drafts and homework assignments!
- If you are unclear about expectations, ask your instructor.

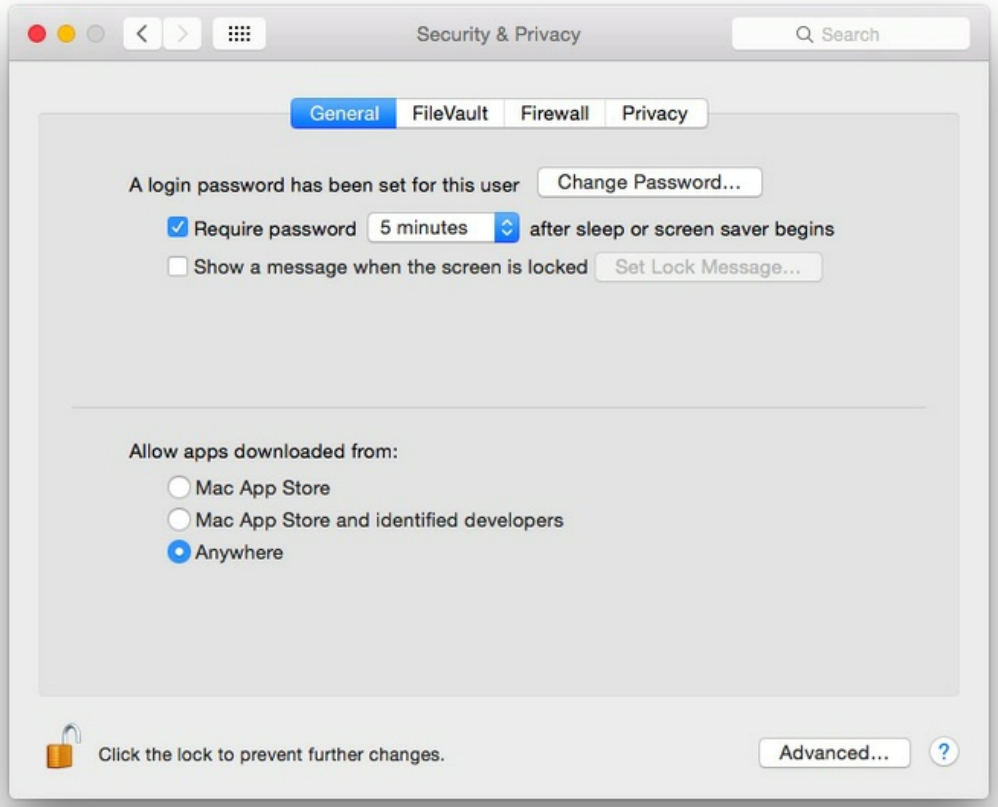
Disability accommodations

If you need disability-related accommodations in this class, if you have emergency medical information you wish to share with the instructor, or if you need special arrangements in case the building must be evacuated, please inform the instructor as soon as possible by seeing him after class or making an appointment to visit during office hours. If you are not currently listed with DSP

(Disabled Students' Program) but believe that you could benefit from their support, you may apply online.

Software and other tools

To install some of the software below, you will need to alter your Mac security preferences to allow apps authored by developers that are not part of Apple's ecosystem. Open `System Preferences>Security & Privacy`, select the `General` tab and click the lock icon at bottom left to make changes, then allow apps downloaded from `Anywhere`:



Firefox

Some of the tools we will use are Firefox add-ons. Ensure that you have the latest version of this popular web browser. Download Firefox, or if already installed, if necessary update to the latest version following these instructions.

Sublime Text or TextWrangler

For some exercises, you will need a text editor optimized for authoring web pages. These are both good options.

ggplot2 web app

Allows you to create charts through a point-and-click interface, unlocking the basic functionality of ggplot2, a visualization package for R (see below). Web app: no installation required.

ColorBrewer

Your go-to resource for color schemes to encode data. Web app: no installation required.

Color Oracle

Allows you to simulate the effects of three forms of colorblindness on your screen, to check that your charts and maps do not exclude the color-impaired. Download and move to your Applications folder. When launched, program icon will appear to the right of your menu bar. You can set to start at login under `Preferences`.

Tableau Public

Visualization software designed for exploratory graphic analysis, which also allows you to create interactive online graphics without JavaScript coding. Download here. You will also need to sign up for a free Tableau Public account.

The Public version of the software requires you to save your visualizations to the open web. However, members of Investigative Reporters and Editors can obtain a license for the Professional version, which allows you to save workbooks on your machine. In any case, I strongly recommend that J-School students join IRE as a student member: it is a leader for training in data journalism.

Table2Clipboard and DownThemAll!

Two handy Firefox add-ons, which will help you extract data from the web. Install them in Firefox here and here.

Kimono

A web scraping tool that works from your web browser. Sign up for a free account from the home page, then install the Bookmarklet. There is also an extension for Google Chrome.

Open Refine

A powerful tool for data cleaning, once owned by Google, now open source. Download the latest

stable version from here. When launched, Open Refine will work in your default web browser; note that your data remains on your computer, and is not uploaded to the web.

Mr. Data Converter

Allows you to convert tabular (spreadsheet/CSV) data to web-friendly formats including JSON. Web app: no installation required.

SQLite Manager

We will manage SQLite databases using this Firefox add-on. Install in Firefox from here. There is no need to separately install SQLite.

GPS Visualizer batch geocoder

Allows you to convert addresses to latitude and longitude coordinates. Web app: no installation required; however, you will need to sign up for API keys for the geocoding services it uses.

CartoDB

A sophisticated cloud-based mapping application that also functions as a geospatial database, allowing you to analyze and process geodata. Sign up here for a free account.

Members of IRE are eligible for an upgraded account with more storage space and additional features.

Leaflet

A JavaScript code library for making interactive online maps. Download the latest stable version from here and unzip, or place the code for the hosted version of this release between the `<head>` `</head>` tags of your HTML code.

QGIS

The leading open source Geographic Information System, which we will use to make maps and process geodata. Download the latest **standard** version of QGIS and its requirements from here. Install GDAL and the Matplotlib Python module before installing QGIS itself.

R and R Studio

R is a software environment and programming language for statistical computing and graphics. Download the latest package for Mavericks and hugher. R Studio provides a user interface that makes it much easier to use. Download the latest installer from here.

D3

A JavaScript code library for manipulating documents based on data, which has emerged in recent years as a powerful and flexible tool for interactive data visualization. Download the latest version from here and unzip, or include the following code snippet in your HTML:

```
<script src="http://d3js.org/d3.v3.min.js" charset="utf-8"></script>
```