COMM 884: Data Visualization for Research

Professor: Sandra González-Bailón
sgonzalezbailon@asc.upenn.edu
Office hours (room 308): Thursday 15:30-16:30

Spring 2016
Thu, 13:30-15:30
Room 225

Course Description

Empirical research employs data to gain insights and build a theoretical understanding of the world. An appropriate visualization of data is key to illuminating hidden patterns and effectively communicate the main findings of research. This course will discuss the visualization strategies of published research, give recommendations of best practice, and discuss tips and techniques for specific research purposes (i.e. hypothesis testing, group comparison) and data structures, including temporal, geographic, and network data. This course will equip you with tools you can use to learn through visualization and to communicate more effectively your own research.

Objectives

*The world is complex, dynamic, multidimensional; the paper is static, flat. How are we to represent the rich visual world of experience and measurement on mere flatland?*

Edward R. Tufte, *Envisioning Information*

Visualization plays different roles in research: it helps you understand your data*, explore its complexities, and uncover patterns that you would have overseen otherwise. Visualization is not tied to quantitative techniques, but it relies on being able to quantify things – for starters, you need to determine the size and resolution of your figures (for which we use inches and pixels). This course will discuss a range of tools you can use to visualize information, from novel plots to correlations; from timelines, to networks. After taking this course, you will know how to:

- Choose the right visualization technique for your data;
- Interpret plots and figures;
- Identify misleading visualizations;
- Communicate your research to a wide audience;
- Find the most persuasive way to convey your research findings.

The course will consist on class discussions where we will consider both the theory and the practice of graphical communication; and a more hands-on application of the concepts discussed via handouts and assignments.

*Data is, in fact, a plural noun (the singular is datum). However it is becoming standard to refer to it as singular, and this syllabus will follow the same convention.*
Assessment

The final grade will be based on three components:

- Attendance and participation (including canvas discussions): 20%
- Five formative assignments: 50%
- Final project: 30%

**Attendance and participation**: the discussion of the readings is central to attain the goals of this course. You will get the 20 points of this component if you consistently participate in class, and you contribute substantive comments and questions that reflect you have carefully read and assessed the material under consideration. Every week there will be a discussion thread in canvas where you have to submit a reaction to the readings. The readings every week will cover (a) visualization theory; and (b) applications in published research. Your weekly submission (no more than one page) should summarize (a) and discuss (b) through the lens of three questions: 1. What is the purpose of the visualization? 2. What kind of data or information is displayed? 3. Could the figure be improved and, if so, how? If there is more than one figure in the reading, you should answer questions 1-3 for each of the figures.

**Formative assignments**: during the course there will be five assignments that will need to be submitted by the deadlines in the schedule below. These assignments are intended for you to show that, in addition to discussing the visualizations created by others, you can also produce visualizations of your own. These assignments will be completed in teams of two. The assignments will require you to work through the handouts distributed every week and discuss, with your team partner, the choices made to complete the assignment. Each assignment will be weighted equally. You will get 10 points in each if you perform the tasks required correctly.

**Final project**: the final project will consist on an academic poster displaying your research (or the research presented in a published article of your choice). Academic posters are one of the most effective ways to present research in conferences – and creating posters forces you to identify the essence (and punchline) of your work. Great posters have few words (~500) and exciting visualizations with good insights into the main question you are trying to answer. The deadline to submit your poster is week 14 (see schedule below). We will discuss all posters in week 15 (last session of the course). A selection of the best posters will be displayed as an online portfolio on the School’s site – it will be a great way to showcase what ASC students do!

The total number of percentage points accumulated with these three components will be converted to the following scale:
A+  |  97-100  
A   |  94-97   
A-  |  90-93   
B+  |  87-89   
B   |  84-86   
B-  |  81-83   
C+  |  78-80   
C   |  74-77   
C-  |  70-73   
D+  |  67-70   
D   |  60-66   
D-  |  0-59    

**General Readings**

Three books are particularly important for this course:


The following books will also be used at different points of this course:

## Schedule

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Sessions</th>
<th>Topics</th>
<th>Deadlines*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January 14</td>
<td>• Class: Introduction to Course</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Introduction to Software</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>January 21</td>
<td>• Class: Telling Stories with Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Graphical Parameters &amp; Packages</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>January 28</td>
<td>• Class: Visualizing Time</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Timelines and time series</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>February 4</td>
<td>• Class: Scales and Proportions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Size, Shape and Scaling</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>February 11</td>
<td>• Class: Designing for an Audience</td>
<td>A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(guest lecture by Bill Beutler, Beutler Ink)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Infographics (templates)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>February 18</td>
<td>• Class: Spatial Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Maps and Spatial Distributions</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>February 25</td>
<td>No class</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>March 3</td>
<td>• Class: Networks</td>
<td>A3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Relational Data</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>March 10</td>
<td>• Class: Trees and Hierarchical Structures</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Treemaps</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>March 17</td>
<td>• Class: Visualizing Text</td>
<td>A4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Semantic Clouds, Nets, and Trees</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>March 24</td>
<td>• Class: Associations and Differences</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Plotting Correlations</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>March 31</td>
<td>• Class: Depicting Error</td>
<td>A5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Data Quality and Missing Values</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>April 7</td>
<td>• Class: Improving Data Displays</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Academic Posters</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>April 14</td>
<td>• Class: When Form Violates Function</td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Handout: Customizing Graphs (Symbols, Text, etc)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>April 21</td>
<td>Poster Presentations</td>
<td></td>
</tr>
</tbody>
</table>

* Legend: A stands for assignment; FP stands for final project (poster). Deadlines apply to Thursday 5pm of the corresponding week.
Weakly Readings

Note: the reading assignments might alter throughout the semester; pay close attention to Canvas for changes and updates. Readings marked with an asterisk are the articles with the graphics you need to assess to answer the three questions: 1. What is the purpose of the visualization? 2. What kind of data or information is displayed? 3. Could the figure be improved and, if so, how?

Week 1 -- Introduction

No background reading. After the course introduction we will discuss these charts and visualizations:


Week 2 -- Telling Stories with Data

Readings:


Week 3 – Visualizing Time

Readings:

- Wainer 2013. *Graphic Discovery*, Chapter 16: There they Go Again!
- Yau, 2011. *Visualize This*, Chapter 4: Visualizing Patterns over Time.

Week 4 – Scales and Proportions

Readings:


Week 5 – Designing for an Audience

Guest lecture by Bill Beutler, BeutlerInk (www.beutlerink.com)

Readings:


Week 6 – Spatial Data

Readings:


Week 7 NO CLASS

Week 8 – Networks

Readings:


**Week 9 – Trees and Hierarchical Structures**

Readings:


**Week 10 – Visualizing Text**

Readings:

- Krum 2013. *Cool Infographics*, Chapter 4: Infographic Resumes

* Lin, Yu-Ru, Drew Margolin, and David Lazer. 2015. "Uncovering social semantics from textual traces: A theory-driven approach and evidence from public statements of U.S. Members of Congress." *Journal of the Association for Information Science and Technology*

**Week 11 – Associations and Differences**

Readings:


**Week 12 – Depicting Error**

Readings:


**Week 13 – Improving Data Displays**


**Week 14 – When Form Violates Function**


**Week 15 – Poster Presentations**

This will be a three hour class –so make arrangements to stay later than usual!