

Project Ideas

Project 1 (Graphic Design) and Project 2 (Interaction Design)

Choose a **dataset to explore and summarize graphically**. You are welcome to use data from your own field or research projects, or consider one of the ideas below.

Statistics MSP students: for at least one Project (1 and/or 2), I encourage you to explore the tract-level U.S. Census Bureau Planning Database. It will be a project for the MSPs in another course this fall, so this will get you a head start. However, you are not required to use this dataset if you would prefer another project.

- U.S. Census Bureau Planning Database:

http://www.census.gov/research/data/planning_database/2015/

How can we understand and predict census mail return rates (dataset variable name `Mail_Return_Rate_CEN_2010`) in each tract, using the other variables? What demographic, socioeconomic, and housing variables are associated with census mail return rates? Are there interesting spatial patterns?

You don't need to build a predictive model (although if you do, find a way to summarize it graphically too). Just show any interesting associations and interactions that you find while exploring the data.

It's OK to draw inspiration from the related Kaggle contest, but of course the work you submit must be your own:

<https://www.kaggle.com/c/us-census-challenge/prospector>

- Haiti 2010 earthquake impact:

http://community.amstat.org/statisticswithoutborders/swb_projects/currentprojects/haiti1

<http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2013.00649.x/abstract>

The organization Statistics Without Borders conducted a survey five months after the January 2010 earthquake in Haiti. How were people's lives affected by the quake? For example, where were most buildings destroyed? How many people relocated nearby vs. far away from their pre-quake residence? How many were still living in temporary shelter five months afterward the destruction? How many households stayed intact vs. became separated? Do these patterns differ for households with different levels of education or employment histories?

I have posted an R version of this dataset as `haiti_pecps.Rdata` but you will still need the codebook and documentation at the SWB link.

Please note that some variables have 2 kinds of missing data:

- “Missing / Don’t Know”: this question makes sense to ask that respondent, but they did not answer. Encoded as NA for numerical data in R dataset.
 - “Not In Universe”: this question does not make sense to ask that respondent. For example, if there is only 1 person in the household, it makes no sense to ask how many cell phones belong to the 2nd person in household. Encoded as NaN for numerical data in R dataset.
- ASA Data Expo: Every few years the American Statistical Association hosts a data visualization contest. Each of the past datasets comes with some questions to explore: <http://stat-computing.org/dataexpo/>
It’s OK to draw inspiration from the contest winners, but of course the work you submit must be your own.
 - FiveThirtyEight: This data journalism site offers many of their datasets on Github. <https://github.com/fivethirtyeight/data>
 - Gapminder: Hans Rosling’s site hosts data (and visualization tools) on global development. How have health care, education, economy, etc. changed in the world’s countries and regions over time?
<http://www.gapminder.org/data/>
 - Sports statistics: Many sports websites provide years or decades of data about athletes, teams, contests, etc. What attributes are associated with success in your favorite sport? The Major League Data Challenge is a contest for visualizing baseball data, with submissions due by Nov 6, 2015:
<http://www.majorleaguedatachallenge.com/>
 - Allen Downey at Olin College has several lists of project ideas:
<https://sites.google.com/site/probstatolin/project>
<https://sites.google.com/site/datascience14/project/descriptions>
<https://sites.google.com/site/datascience15/project/descriptions>

Project 3 (Research)

Choose **an academic research paper/presentation** (or several related papers on one topic) to summarize in a report.

In your report, you will take the authors’ proposed solution to a graphical problem and use it to plot a new dataset. So, the paper(s) you read should propose a concrete solution to a problem in data visualization: a specific way to display uncertainty, to visualize models, to interact with graphics, etc. (Do not choose a very abstract paper about, say, a new framework for classifying graphs, because there will be nothing specific for you to reproduce.)

You are welcome to use anything of interest, or consider one of these ideas:

- Displaying uncertainty in general:
 - Correll & Gleicher (2014), “Error Bars Considered Harmful”
<https://graphics.cs.wisc.edu/Papers/2014/CG14/Preprint.pdf>
 - Wood et al. (2012), “Sketchy Rendering for Information Visualization”
http://openaccess.city.ac.uk/1274/1/wood_sketchy_2012.pdf
 - Andy Kirk, blog post with more references
<http://www.visualisingdata.com/2015/02/references-visualising-uncertainty/>
- Displaying uncertainty on maps:
 - MacEachren et al. (2012), “Visual Semiotics & Uncertainty Visualization”
http://www.geovista.psu.edu/publications/2012/MacEachren_IEEE_TVCG_PrePub_2012_reduced_res.pdf
 - Francis et al. (2012), “Alternative Strategies for Mapping ACS Estimates and Error of Estimation”
https://pad.human.cornell.edu/papers/downloads/CPC_Strategies%20for%20Mapping%20ACS%20Estimate%20and%20MOE.pdf
 - Griffin (2015), “Visualizing Attribute Uncertainty in the ACS”
<http://www.ncrn.info/event/ncrn-virtual-seminar-feb-4-2015>
- Displaying multiple comparisons:
 - Almond et al. (2000), “Displays for Comparing a Given State to Many Others”
<http://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474517>
 - Wright et al. (2014), “Ranking Populations Based on Sample Survey Data”
<https://www.census.gov/srd/papers/pdf/rrs2014-12.pdf>
- Displaying missing data:
 - Cook & Swayne (2007), “Interactive and Dynamic Graphics for Data Analysis,” Ch. 3
<http://www.ggobi.org/book/>
<http://link.springer.com/book/10.1007/978-0-387-71762-3>
 - Templ et al. (2012), “Exploring incomplete data using visualization techniques”
<http://feb.kuleuven.be/public/n11027/publications/ADAC2012.pdf>
 - Prantner (2011), “Visualization of imputed values using the R-package VIM”
<https://cran.r-project.org/web/packages/VIMGUI/vignettes/VIM-Imputation.pdf>
- Displaying survey-weighted data:
 - Korn & Graubard (1998), “Scatterplots with Survey Data”
<http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1998.10480541>
- Displaying network data:

- Krzywinski et al. (2012), “Hive plots—rational approach to visualizing networks”
<http://bib.oxfordjournals.org/content/13/5/627>
- Heer & Boyd (2005), “Vizster: Visualizing Online Social Networks”
<http://vis.berkeley.edu/papers/vizster/2005-Vizster-InfoVis.pdf>
- Displaying ratings data:
 - Robbins & Heiberger (2011), “Plotting Likert and Other Rating Scales”
http://www.amstat.org/sections/srms/proceedings/y2011/files/300784_64164.pdf
 - Heiberger & Robbins (2014), “Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications”
<http://www.jstatsoft.org/article/view/v057i05>
- Visual diagnostics for statistical/ML models:
 - Hadley Wickham, blog post with several papers and R packages
<http://had.co.nz/model-vis/>
- Visual statistical inference:
 - Buja et al. (2009), “Statistical inference for exploratory data analysis and model diagnostics”
<http://rsta.royalsocietypublishing.org/content/367/1906/4361.short>
- Collaborative and narrative visualization:
 - Heer et al. (2009), “Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization”
<http://vis.stanford.edu/papers/senseus>
 - Segel & Heer (2010), “Narrative Visualization: Telling Stories with Data”
<http://vis.stanford.edu/papers/narrative>
- Visual attention:
 - Kim & Varshney (2008), “Persuading Visual Attention through Geometry”
https://www.cs.umd.edu/gvil/papers/kim_TVCG08.pdf
 - Ooms et al. (2014), “Study of the attentive behavior of novice and expert map users using eye tracking”
<http://www.tandfonline.com/doi/abs/10.1080/15230406.2013.860255>
- Color:
 - Eddins (2014), “Rainbow Color Map Critiques”
<http://blogs.mathworks.com/steve/2014/10/20/a-new-colormap-for-matlab-part-2-troubles-with-rainbows/>
“A New Colormap for MATLAB”
<http://www.mathworks.com/company/newsletters/articles/rainbow-color-map-critiques-an-overview-and-annotated-bibliography.html>

- Setlur & Stone (2015), “A Linguistic Approach to Categorical Color Assignment for Data Visualization”

<https://research.tableau.com/paper/linguistic-approach-categorical-color-assignment-data-visualization>

- Robert Kosara, blog posts on recent conference presentations:

- IEEE VIS 2014

<https://eagereyes.org/blog/2014/vis-2014-observations-and-thoughts>

- CHI 2015

<https://eagereyes.org/blog/2015/conference-report-chi-2015>

- EuroVis 2015

<https://eagereyes.org/blog/2015/report-eurovis-2015>