Week 15

Text

INFO 3402: Information Exposition

Brian Keegan, Ph.D.
Assistant Professor, Information Science
brian.keegan@colorado.edu

University of Colorado Boulder

Course Overview

Module	Week	Dates	Type	Skill
Shaping	1	Jan 11, Jan 13	Computation	Loading
	2	Jan 18, Jan 20	Computation	Aggregating
	3	Jan 25, Jan 27	Computation	Joining
	4	Feb 1, Feb 3	Computation	Tidying
Distribution	5	Feb 8, Feb 10	Computation	Histograms
	6	Feb 15, Feb 17	Communication	Audience
Comparison	7	Feb 22, Feb 24	Computation	Cat plots
	8	Mar 1, Mar 3	Communication	Persuasion
Trend	9	Mar 8, Mar 10	Computation	Time series
	10	Mar 15, Mar 17	Communication	Uncertainty
	11	Mar 22, Mar 24	Spring Break	
Relationship	12	Mar 29, Mar 31	Computation	Scatter plots
	13	Apr 5, Apr 7	Communication	Fallacies
Spatial	14	Apr 12, Apr 14	Computation	Choropleths
	15	Apr 19, Apr 21	Communication	Conventions
Projects	16	Apr 26, Apr 28	Projects	

Text

Readings

○ Questions for Friday's Weekly Quiz 15 will be drawn from these readings

  ○ Pandas User Guide – Working with text data

  ○ Schwabish (2021) – Ch. 10: Qualitative

  ○ Dykes (2019) - Ch. 6, 7, & 8: The Foundation, Setting and Polishing the Scenes of Your Data Story

Weekly Assignment 15

○ Submit an outline of your final project

○ Follow the practices in the Dykes chapters

- ○ Identify a narrative model and/or storytelling arc (Ch 6)
- ○ Storyboard the data, identify heroes and conflict (Ch 6)
- ○ Choose the right kinds of visualizations (Ch 7)
- ○ Grouping or removing less important values (Ch 7)
- ○ Remove distractions, reflect on labeling choices (Ch 8)
- ○ Analogies and examples that can be used? (Ch 8)

○ Doesn't have to follow this exact structure, but needs to have outline of story and sketches

- ○ Submit as a PDF of a slides (easy to draft) or photos of hand-drawn outline

Module Assignment 06

Module Assignment 06

- Difficulties in installing geospatial libraries for everyone → we're punting on it all
  - Submit a Module Assignment using geospatial or text tools if you want to, otherwise:

- Revise and resubmit a previous module assignment for a regrade
  - Your previous grade will still stand

Final Project

Final Project

○ Write a data-driven op-ed using a combination of visualization methods

    ○ At least three different visualization methods (histograms, catplots, time series, scatter plots, choropleths)

    ○ Can use any of the datasets or sources from class or your own data

    ○ Can revisit and expand on a previous module assignment

    ○ Motivation, structure, persuasion, causal reasoning, following conventions must be very strong!

○ >1000 words with at least **three** visualizations

○ Final Project will be due on **<u>Wednesday, May 4 by 11:59pm</u>**

    ○ Submit URL of your Medium post to Canvas or save and submit as an HTML file

    ○ Tag your post on Medium with "INFO3402S22FP" and whatever other tags you'd like

Notebook

Notebook

- Download "Week 15 – Lecture.ipynb" and "politician_tweets.csv"

- Working with Python's NLTK library and text methods in pandas
- Text processing: retrieval, tokenizing, filtering, normalizing
- Working with a single large documents or a collection of small documents
- Sentiment analysis

Next class

Next Class

- Review concepts and exercises from last class
  - Complete "Thursday Questions" form! https://forms.gle/YA8RBDw4cRu3XWgv7 (ungraded/optional)

- Time to brainstorm and work on Weekly Assignment 14 and Module Assignment 06

- Weekly quiz at the end of class (12:00–12:30)

- Upcoming deadlines
  - Weekly Assignment 14 due Sunday, April 17 before midnight
  - Module Assignment 05 due Wednesday, April 13 before midnight
  - Module Assignment 06 due Wednesday, April 27 before midnight