Week 14

Choropleths

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	

Readings

- Questions for Friday's Weekly Quiz 13 will be drawn from these readings
  - Wilke 2019 – Visualizing Geospatial Data
  - Schwabish 2021 – Ch 9: Geospatial
  - Introduction to GeoPandas

Weekly Assignment 14

- **Skills**: Visualizing spatial data using GeoPandas

- **Data**: Colorado elections

Module Assignment 05

Module Assignment 05

○ Use the U.S. county data to identify an *unusual* relationship between variables

- ○ ~1,000 different county-level variables in "us_counties.csv", "analytic_data2021.csv", "Unemployment.csv"
- ○ Must be unusual: A relationship between poverty and income is **not** unusual—it's to be expected!
- ○ Some relationships might be trivial: employment from one dataset is likely similar with employment from another
- ○ The relationship should be strong-ish: if it's linear, a correlation above 0.2 or below -0.2
- ○ Identifying this relationship can be top-down (sorting, correlograms, pairplots, *etc*.) or bottom-up (exploring pairs)
- ○ Like WA12 Questions 2 & 3, explore whether this relationship shows up or disappears in related variables
- ○ Make a case for there being a causal rather than a random relationship between these variables

○ 700 – 1000 words with at least one visualization

○ Module Assignment 05 will be due on **Wednesday, April 13 by 11:59pm**

- ○ Submit URL of your Medium post to Canvas or save and submit as an HTML file
- ○ Tag your post on Medium with "INFO3402S22A5" and whatever other tags you'd like

Module Assignment 3 shout-outs

- Jessie Bart – Worldwide Income Inequality
  - Exploring changing in Gini coefficient of income inequality over time and by country
- Max Blanco – World War II's Effects on Population Growth
  - Comparing demographics of US, Germany, and Russia
- Chris Davis – Free and Independent Media Trends
  - Contextualizing Russian media propaganda with Democracy Index and Independent media scores
- Bailey Gimpel – A Humanitarian Crisis: Civil Liberties on the Decline
  - Identifying interesting tension between declining civil liberties scores but increasing political participation
- Jasmine Rivera – Birthrates declining at an Alarming Rate in Japan
  - Nice demographic pyramids, motivation from a 2017 article, comparing different kinds of evidence
- John Ross Greene – The Growing Inequality of Wealth and Power in the United States
  - Visualization stacked with lots of accessibly information, strong voice throughout writeup
- Brian Lee – South Korea retiring into Economic Collapse
  - Nice demographic pyramids and explanations, strong motivation and voice, triangulating evidence
- Max Vali – Income, Life Expectancy, and Access to Sanitation
  - Exploring trends and interactions across three variables over time

Module Assignment 06

Module Assignment 06

○ Use the U.S. county data to visualize interesting spatial variation of **two** features

   ○ "interesting spatial variation of a feature" → the counties need to change in an interesting way

   ○ Motivate why you're looking at the spatial variation of these features: urban-rural, north-south, *etc*.

   ○ Make two choropleths: can be different features, an aggregation of county to state, something else

   ○ Are there similar or different spatial patterns in both features?

   ○ Are there patterns in the spatial variation that is missed by a scatterplot, state-level vs. county-level?

○ 700 – 1000 words with at least **two** visualizations

○ Module Assignment 06 will be due on **<u>Wednesday, April 27 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 "INFO3402S22A6" and whatever other tags you'd like

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 **two** visualizations

- Module Assignment 06 will be due on **Wednesday, May 4 by 11:59pm**
  - 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 "us_states.zip", "boco_precincts.zip", "1976-2020-president.csv", and "Week 14 – Lecture.ipynb"

- Working with Python's spatial libraries like geopandas, geoplot, cartopy, *etc*.

- Using pandas functionality with geopandas

- Working with map projections

- Making and customizing choropleth visualizations

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