Week 07

Categorical plots

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	Aggregation
	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	Trust
	11	Mar 22, Mar 24	Spring Break	
Relationship	12	Mar 29, Mar 31	Computation	Scatter plots
	13	Apr 5, Apr 7	Communication	Annotation
Spatial	14	Apr 12, Apr 14	Computation	Choropleths
	15	Apr 19, Apr 21	Communication	Conventions
Projects	16	Apr 26, Apr 28	Projects	

Shift in course structure

- We've been focusing on the computational side and slighting the communication side
  - Module Assignment 01 helped us calibrate where class needs are

- This week will have a stronger focus on the communication side (with some computation)
  - Weekly assignment will also support framing, analyzing, and writing up your MA02

- The first week of a module will be more computational, the second week more communication

- No changes in pace of weekly assignments, weekly quizzes, or module assignments

Weekly overview

○ Notebook exercises (ungraded) – Assigned Tuesdays and reviewed Thursdays

○ Add a Visualization Critique (ungraded) – Discussed in class on Thursdays

○ Complete "Thursday Questions" form! (ungraded/optional) – Review in class on Thursdays

○ Weekly Assignment (graded, 2%) – Assigned Tuesdays and due Sundays by midnight

○ Weekly Quiz (graded, 1%) – End of class on Thursdays (open 12:00pm – 12:30pm)

○ Module Assignment 02 (graded, 5%) – Due Tuesday, February 22 by 11am*

* We'll discuss due dates in a few slides

Readings

○ Questions for Friday's Weekly Quiz 07 will be drawn from these readings

○ It's important to be exposed to many examples of data visualization and storytelling!

○ Seaborn documentation on plotting with categorical data

   ○ How to translate tidy data into different categorical plots, included using hues

   ○ Also look at https://www.python-graph-gallery.com/

○ EIU Democracy Index 2021

   ○ Pay attention to how they use, visualize, and narrate their findings to tell a story!

○ Schwabish, J. (2021). Chapter 4: Comparing Categories.

   ○ Best practices for designing visualizations involving comparisons of categories and continuous data

Weekly Assignment 07

- **Skills**: Reviewing data cleaning and plotting concepts, categorical plots

- **Data**: EIU Democracy Index

Module Assignments 02 and 03

Module Assignment 01 feedback

- Write for a general audience, not to me and Samantha

  - Open with a question, provocation, case …anything to engage the reader

  - Do not assume the reader knows anything about data analysis and visualization, programming, *etc*.

  - "Like you said in lecture, I did X"

- There is a place for tutorial-style writing on Medium, but that's not what we're focused on

  - "I used X_df and the Y method to turn this into Z which looks like this"

  - We don't want a step-by-step narration of your analysis; we want narratives with polished results

- Quality matters

  - Use high-quality saved figures (docs) rather than screenshots whenever possible

  - No glaring spelling or grammatical problems; professional tone and style

Module Assignment 02

- Use the House expenditure data to tell a **story** about a US House disbursement data

  - You're welcome to bring in other data if it helps with your story, but disbursement data at a minimum

  - Emphasis on **story**! Engage the reader, identify characters/issues, build and resolve tension

  - Use WA06 to think about the audience you want to write for

- 700 – 1000 words with at least one visualization (extra credit for visualizations with annotation!)

  - Mine is posted already: https://medium.com/information-expositions-s2022/congressional-disbursements-by-party-and-incumbency-5aec7420b7eb

- Module Assignment 02 will be due on **Wednesday, February 23 by 11am**

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

  - Tag your post on Medium with "INFO3402S22A2" and whatever other tags you'd like

Module Assignment 03

- Use the World Development Indicators, Democracy Index, or a similar country-level longitudinal dataset to tell a **story** about how countries, regions, and/or the world are changing
  - You're welcome to bring in other data if it helps with your story
  - Emphasis on **story**! Engage the reader, identify characters/issues, build and resolve tension
  - Use WA08 to think about the audience you want to write for

- 700 – 1000 words with at least one visualization (extra credit for visualizations with annotation!)

- Module Assignment 03 will be due on **Wednesday, March 9 by 11am**
  - Submit URL of your Medium post to Canvas or save and submit as an HTML file
  - Tag your post on Medium with "INFO3402S22A3" and whatever other tags you'd like

Democracy Index and World Development Indicators

Democracy Index

- *The Economist* (the magazine) has a sister company that publishes data and forecasts, including their most famous data product: the Democracy Index
  - Annual reports from 2006, 2008, and 2010 – 2021
  - Experts' assessments of 60 questions in five categories about political institutions and freedoms
    - Electoral process, civil liberties, functioning government, political participation, and political culture
- Based on average of scores, countries are classified into:
  - full democracies, flawed democracies, hybrid regimes, and authoritarian regimes
  - United States fell from a full to a flawed democracy in 2016 and score has continued to decline
- Similar scores and rankings published by many other organizations
- Skim the 2021 report to learn more about the methodology, analyses, and visualizations

World Development Indicators

○ The World Bank collects and publishes country-level longitudinal data for more than 1,400 different types of social and economic data

   ○ https://databank.worldbank.org/source/world-development-indicators

   ○ Select countries (or checkbox for all) > select series (probably should search) > select time frames > Apply Changes > Download options > CSV

○ I've pulled out some classic indicators like GDP and labor population as well as some numbers specific to military expenditures and size

Exploratory questions

Exploratory questions need to be refined to more specific analyses, audiences, motivations!

Category plots

Notebook time!

○ Download "Week 07 – Lecture.ipynb"

　○ Previous data about Congressional disbursements and new data about countries' military expenditures

Next class

Next Class

○ Review concepts and exercises from last class

　　○ Complete "Thursday Questions" form! https://forms.gle/DBjD4FpXVpsP89dv6　(ungraded/optional)

○ Critique a data narrative or visualization

　　○ Share a link & a few sentences about a data story or visualization (ungraded/optional)

○ Time to work on Weekly Assignment

　　○ Weekly Assignment due on Sunday by midnight by submitting HTML notebook to Canvas

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