CSC/SDS 109: Communicating with Data

HW 02: First Tableau Visualizations

This is an individual or pair assignment-- you pick (collaboration is highly encouraged)!

Goals:

- Work with real data
- Use Tableau to visualize data

Instructions

Step 1: Obtain Data

- Choose one of the datasets under the "In-class" tab on the course website:
    - **Use the dataset on the course website – they've been cleaned for you**
    - College.csv (source: https://www.kaggle.com/datasets/yashgpt/us-college-data)
    - cereal.csv (source: https://perso.telecom-paristech.fr/eagan/class/igr204/datasets)
    - palmerpenguins.csv (source: https://allisonhorst.github.io/palmerpenguins/articles/intro.html)
    - bluebikes-tripdata_sm.xlsx (source: https://data.boston.gov/dataset/blue-bikes-system-data)
- Load the dataset into Tableau and explore the dimensions. For example, if you download the cereal dataset, you'll see something like this:

Name	Mfr	Type	Calories	Protein	Fat	Sodium	Fiber	Carbo	Sugars	Potass
100% Bran	N	C	70	4	1	130	10.0000	5.0000	6	280
100% Natural Bran	Q	C	120	3	5	15	2.0000	8.0000	8	135
All-Bran	K	C	70	4	1	260	9.0000	7.0000	5	320
All-Bran with Extra Fiber	K	C	50	4	0	140	14.0000	8.0000	0	330
Almond Delight	R	C	110	2	2	200	1.0000	14.0000	8	-1
Apple Cinnamon Cheerios	G	C	110	2	2	180	1.5000	10.5000	10	70
Apple Jacks	K	C	110	2	0	125	1.0000	11.0000	14	30
Basic 4	G	C	130	3	2	210	2.0000	18.0000	8	100
Bran Chex	R	C	90	2	1	200	4.0000	15.0000	6	125
Bran Flakes	P	C	90	3	0	210	5.0000	13.0000	5	190

cereal.csv — 16 fields 77 rows

- Review your dataset. Answer the following:
    - Who collected the data? (If you can't tell, record how you searched for this information.)
    - When was the data collected? (If you can't tell, record how you searched for this information.)
    - What does each observation (row) represent?
    - What variables (columns) are included?
    - Is the dataset representative of the population it seeks to capture? (ex. If the dataset claims to be US Colleges, are college across the entire country included? Are public and private institutions included? Etc.)
    - Pick one of the questions you answered above. What implications does your answer have for biases or ethical issues present in the data?
- Using Tableau, create three DIFFERENT visualizations that each show something interesting in your data. Each visualization must:
    - Show unique dimensions of the dataset (i.e. not the same as your other two visualizations)
    - Be a unique visual encoding (i.e. not the same as your other two visualizations)
    - Include:
        - Descriptive title
        - Readable axis titles
        - Readable axis labels
        - A legend when necessary
        - Zero double encodings
        - Appropriate data-visual mappings
- Take a screenshot of your completed visualizations and add them to the document you will turn in.
- Below each visualization briefly describe what interesting trend is shown.

Submission

Submit your deliverable(s) as a PDF on Gradescope. If you worked with a partner, submit as a group (https://guides.gradescope.com/hc/en-us/articles/21863861823373-Adding-Group-Members-to-a-Submission).

Rubric

The following matches the rubric you will see on Gradescope.

	Points	Criteria
General:	0.5	Data collector (or search procedure)
	0.5	Data collection time (or search procedure)
	0.5	ID observation
	0.5	ID variables
	0.5	Representativeness of data
	0.5	Data biases
For each visualization:	0.5	Screenshot
	2	Description of interesting trend
	1	Unique dimensions
	1	Unique visual encoding
	0.5	Descriptive title
	0.5	Readable axis titles
	0.5	Readable axis labels
	0.5	A legend when necessary
	0.5	Zero double encodings
	2	Appropriate data-visual mappings
TOTAL	30	