STORYTELLING WITH DATA

Introduction to Data Viz & Tools(Tableau)

RAFIA MIRZA

Humanities Librarian, SMU
Digital Humanities
DHRI coordinator

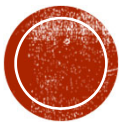
"STORYTELLING WITH DATA: A HANDS-ON WORKSHOP FOR BEGINNERS" OUTLINE

- Your research data tells a story. Interactive data visualization is about communicating your insights and research effectively, giving your data a voice.

- The workshop will cover:
  - Best practices in data visualization
  - Data visualization tools
  - Tableau basics through demonstration and hands-on learning

- Data
  - Question/ Initial Goal
  - Data analysis
  - Data Viz
    - Data storytelling
  - Publication / End Goal

- Tools
  - Tableau(s)
    - Demo
    - Hand on



DISCUSSING DATA

WHERE ARE YOU ON THIS CYCLE? WHICH STEPS APPLY TO YOU?

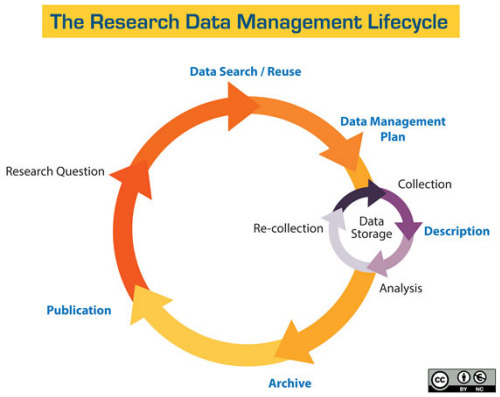


Image via US Santa Cruz Library RDM page

- Research questions
  - Why are you looking for data? What kind of data do you think will have useful information for your questions?
- Data Search/ Reuse
  - You are looking for already existing data or data sets that you can reuse for your research.
  - What is the difference between data and a data set?
  - Is the data you found in the proper format for analysis? Is it clean and structured?
- Data Management Plan (DMP)
  - Do you need a long term plan to manage your data?
  - Do you want to share it with others?
  - Does the grant you are working on require you have a plan for sharing and/or preserving your data?
- Data Storage (Collection, Description, Recollection)
  - Where are you going to store the data?
  - Do you want to store the description/metadata?
- Analysis
  - You are deciding what tools (Excel) or Scripts (Python, R) you are going to use to ask some questions of your data)
- Archive
  - You are done using the data, but you want to save it for the long term.
- Publications
  - You are now publishing your results, research and/or the data set.
  - Data visualizations are a type of publication.

The Data Life Cycle

generation	collection	processing	storage	management	analysis	visualization	interpretation
"People generate data: every search query we perform, link we click, movie we watch, book we read, picture we take, message we send, and place we go contribute to the massive digital footprint we each generate" *[Think also of historical source documents]*	*"Not all data generated is collected, perhaps out of choice because we do not need or want to, or for practical reasons.... Deciding what to collect defines a filter on the data we generate"*	*"everything from data cleaning, data wrangling, and data formatting to data compression, for efficient storage, and data encryption, for secure storage"*	*"the bits are laid down in memory"*	*"We are careful to store our data in ways both to optimize expected access patterns and to provide as much generality as possible....We need to create and use different kinds of metadata for these dimensions of heterogeneity to maximize our ability to access and modify the data for subsequent analysis"*	*"all the computational and statistical techniques for analyzing data for some purpose: the algorithms and methods that underlie artificial intelligence (AI), data mining, machine learning, and statistical inference, be they to gain knowledge or insights, build classifiers and predictors, or infer causality"*	*"helps present results in a clear and simple way that a human can readily understand and visualize"*	*"we provide the human reader an explanation of what the picture means. We tell a story explaining the picture's context, point, implications, and possible ramifications"*

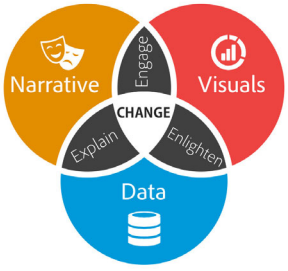
Jeannette M. Wing (2019) "The Data Life Cycle" *Harvard Data Science Review*

Handout via Chris Alen Sula's work, "Nuances of Data: What Can DH Contribute?" Association for Computers & Humanities, Pittsburgh, Penn., 24 July 2019.

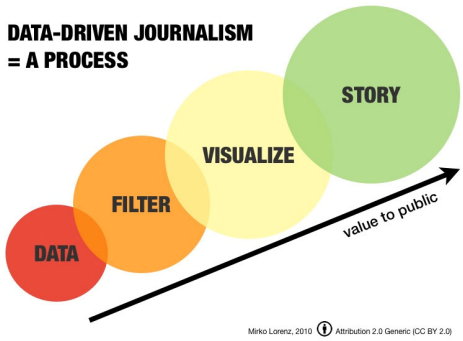
DISCUSSING DATA STORYTELLING

STORY TELLING

What story are you trying to tell?



Who are you trying to tell this story to?



DATA VISUALIZATION CHECKLIST: CHART CHOICE

- What type of chart should you choose?
- What information are you trying to convey with you chart?

- What type of data do you have?
- What type of relationships are you trying to analyze or show?
  - Comparison
  - Distribution
  - Relationship
  - Composition

Chart Choosers

- The Data Visualisation Catalogue
- Depict Data Studio
- Data To Viz
- Data Visualization (in Tableau and Excel)
- Tableau Public
- Which Type of Chart or Graph is Right for You? Tableau

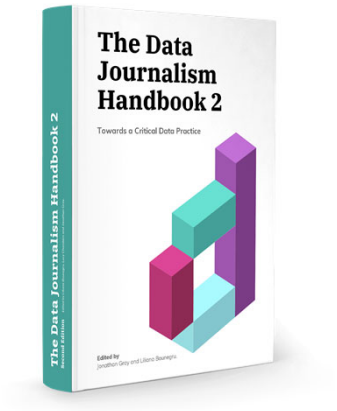
DATA VISUALIZATION CHECKLIST

- Do your choices help your narrative or distract from you narrative?
  - Text/Font
  - Arrangement/Sizing
  - Color
  - Lines

Data Visualization Checklists

- Berkeley BPMO
- Interhacktives
- DataVizChecklist
- Evergreen Data
- Common Caveats

DATA JOURNALISM

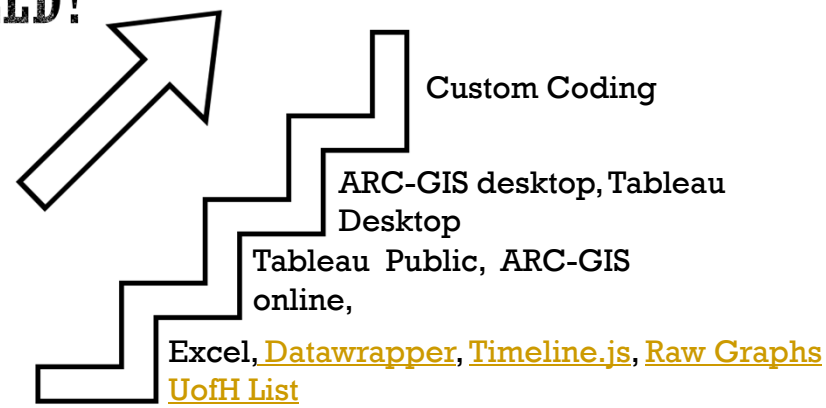


- Datajournalism.com
- Introduction to Data Journalism (UofH Libraries)
- Data Journalism sites
  - The Pudding
  - The Markup
  - FivethirtyEight

DISCUSSING TOOLS

Tableau

HOW ROBUST AND COMPLEX A TOOL DO YOU NEED?



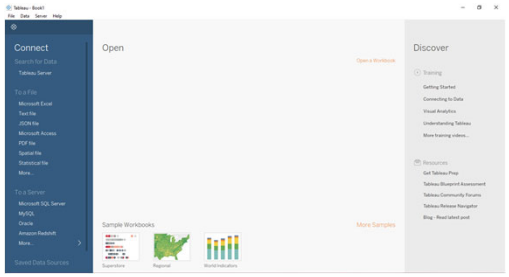
Created by Montu Yadav
from Noun Project

WHAT IS TABLEAU?

- Why choose Tableau?
- **Tableau options**
- Tableau product suite: Tableau Prep, Tableau Desktop, Tableau Server & Tableau Online, Tableau Mobile.
- **Tableau Public**
- "Tableau Public is **free software** that can allow anyone to connect to a spreadsheet or file and create interactive data visualizations **for the web.\"**
- **Tableau Desktop**
- Tableau Desktop is part of the Tableau Creator suite. It allows you to **analyze data** and **create visualizations** on your desktop. (Academic program)

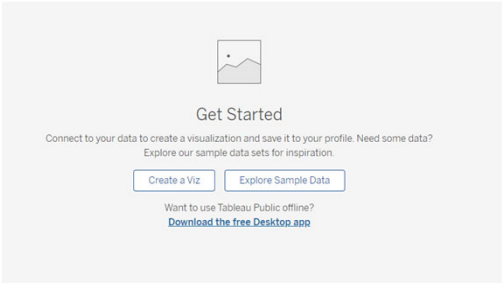
TABLEAU

Desktop



Public

(Can download a desktop application)



HOW TABLEAU WANTS YOUR DATA TO LOOK

Not this

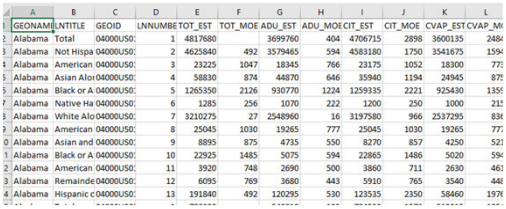
WINE SALES IN THE U.S.

(Wine shipments in millions of 9-liter cases from California, other states and foreign producers entering U.S. distribution)

From WineInstitute.org (http://www.wineinstitute.org/resources/pressroom/05012017)

Year	Table Wine (millions of 9L cases)	Dessert Wine (millions of 9L cases)	Sparkling Wine (millions of 9L cases)	Total Wine (millions of 9L cases)	Total Retail Value ($) (billions)
2016	331.7	41.9	25.6	399.2	59.5
2015	324.9	[illegible]	22.5	387.7	57.1
2014	323.7	[illegible]	20.6	378.8	55.5
2013	327	[illegible]	18.9	377.5	52.3
2012	319.5	[illegible]	17.9	367.7	50.8
2011	308.1	[illegible]	17.5	357	48.6
2010	290.8	[illegible]	15.4	335	46.5
2009	282.4	27.2	14	323.5	45.2
2008	272.2	27.7	13.6	313.5	45
2007	272.5	26.7	13.9	313	43.5

Data source



GEONAME	LNTITLE	GEOID	LNNUMBE	TOT_EST	TOT_MOE	ADU_EST	ADU_MOE	CIT_EST	CIT_MOE	CVAP_EST	CVAP_M
Alabama	Total	04000US0	1	4817680		3695760	404	4706715	2898	3600135	248
Alabama	Not Hispa	04000US0	2	4625840	492	3579465	594	4583180	1750	3541675	159
Alabama	American	04000US0	3	23225	1047	18345	766	23175	1052	18300	77
Alabama	Asian Alor	04000US0	4	58830	874	44870	646	35940	1194	24945	87
Alabama	Black or A	04000US0	5	1265350	2126	930770	1224	1259335	2221	925430	135
Alabama	Native Ha	04000US0	6	1285	256	1070	222	1200	250	1000	21
Alabama	White Alo	04000US0	7	3210275	27	2548960	16	3197580	966	2537295	83
Alabama	American	04000US0	8	25045	1030	19265	777	25045	1030	19265	77
Alabama	Asian and	04000US0	9	8895	875	4735	550	8270	857	4250	52
Alabama	Black or A	04000US0	10	22925	1485	5075	594	22865	1486	5020	59
Alabama	American	04000US0	11	3920	748	2690	500	3860	711	2630	46
Alabama	Remainde	04000US0	12	6095	769	3680	443	5910	765	3540	44
Alabama	Hispanic c	04000US0	13	191840	492	120295	530	123535	2350	58460	197

OPEN APPLICATION & USE SAMPLE DATA

- Dimensions and Measures, Blue and Green
  - Discrete Dimension
    - qualitative values (like names)
  - or
  - Continuous Measure
    - quantitative values (numbers)
- Double click or drag over
  - Dimension define number of marks
  - Measures can be aggregated
    - Sum? Disaggregate?
- Show me!

DATASETS

Get s dataset to analyze

- Tableau – use sample dataset
- Or you can Download a workbook

Demo dataset

- Cat vs Dog Popularity in the US
  - Population and ownership by household of dogs and cats broken down by state via American Veterinary Medical Association.
  - Dataset (xlsx)
- Play around with data or you can try to answer a specific question
- See what others did with the same data

GETTING STARTED!