

DATA VISUALIZATIONS IN ACADEMIC LIBRARIES

Jenny Wong-Welch

STEM Librarian & Director, build IT makerspace

San Diego State University

CSU Lib IT 2019,
Northridge, CA

August 7th, 2019

ME

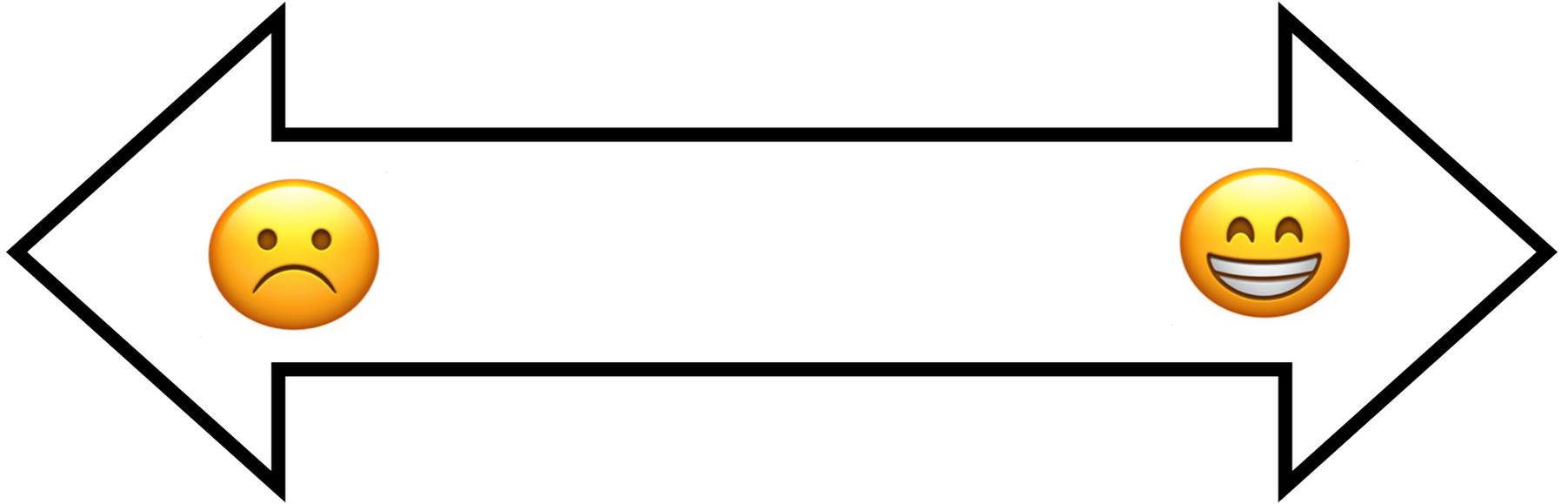
- Crafty
- Libraries are for experimenting
- Background in Math & Economics
- Achiever, Analytical, Deliberative, Focus, Strategic
 - “Strengths Finder 2.0” Tom Rath



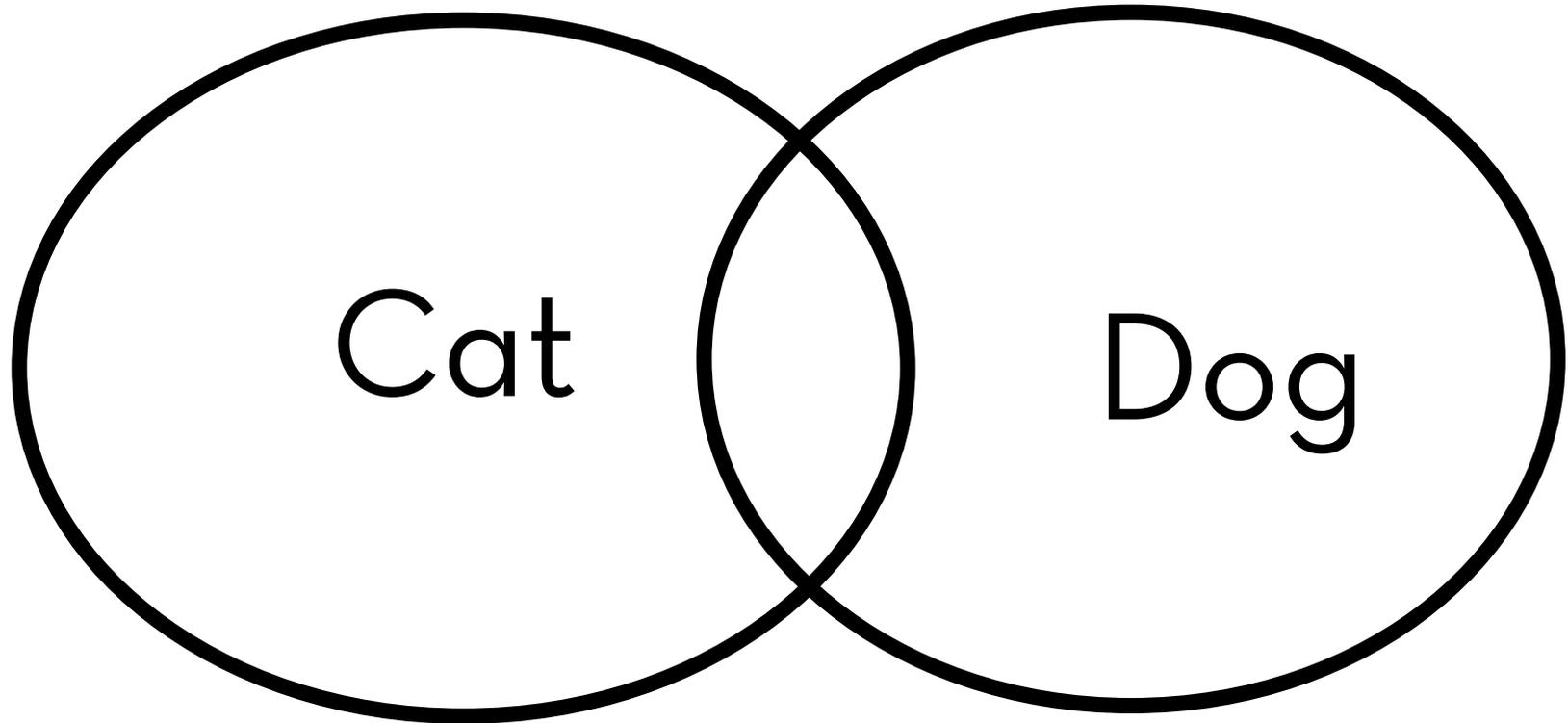
**ARE YOU ORIGINALLY
FROM CALIFORNIA?**

**IS A HOT DOG A
SANDWICH?**

MY COMFORT-LEVEL WITH CREATING DATA VISUALIZATIONS IS

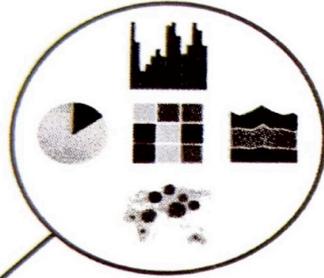


WHICH ANIMAL IS THE BEST?

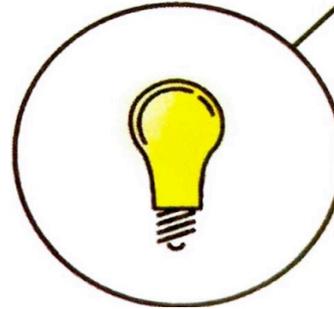
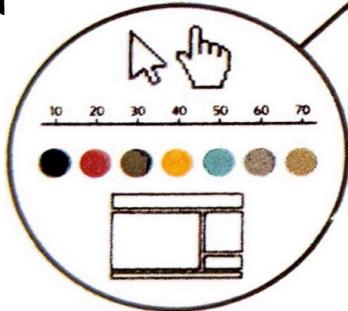


DATA VISUALIZATIONS

Definition



The representation and presentation of data to facilitate understanding



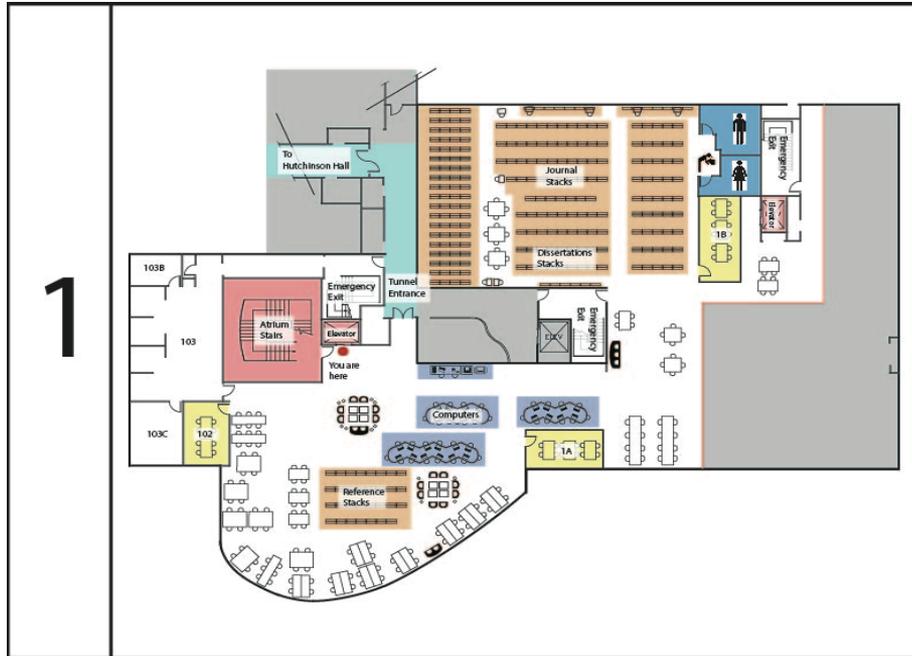
DATA VISUALIZATION

- an umbrella term, covering both information and scientific visualization
- A general way of talking about **anything that converts data sources into a visual representation**
 - like charts, graphs, maps, sometimes even just tables
- **Scientific visualization:** the visualization of scientific data that have close ties to real-world objects with spatial properties
- **Information visualization:** covering most statistical charts and graphs
- **Infographic:** combining various statistics and visualizations with a narrative

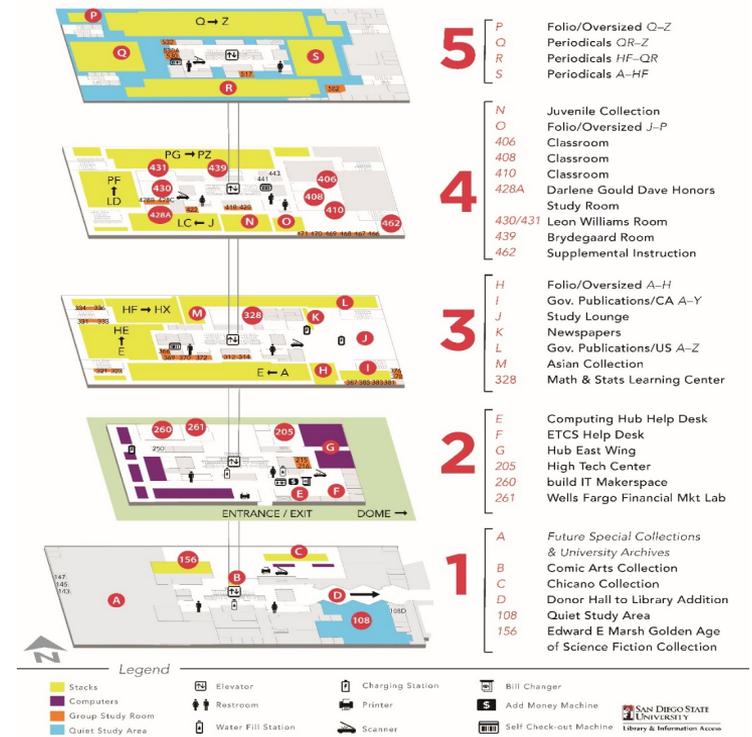
WHY SHOULD WE USE DATA VISUALIZATIONS?

USER SERVICES

LIBRARY MAPS



LOVE LIBRARY (LL)

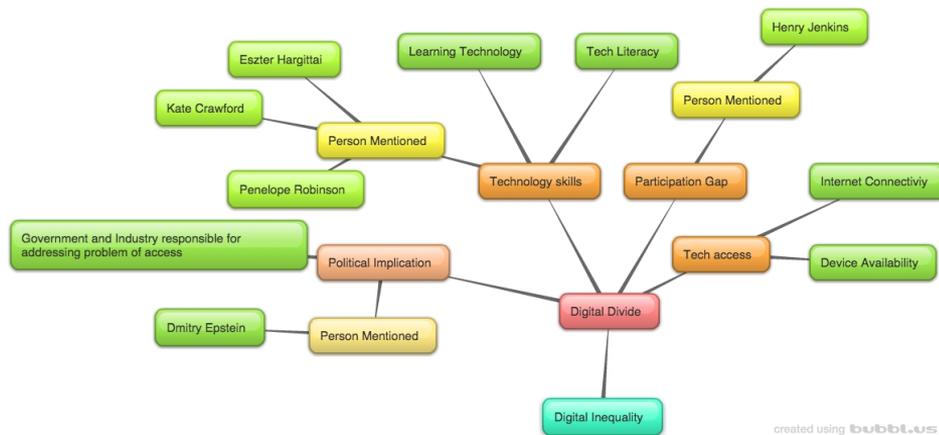


INSTRUCTION

Mind Map of Bullying from Credo Reference

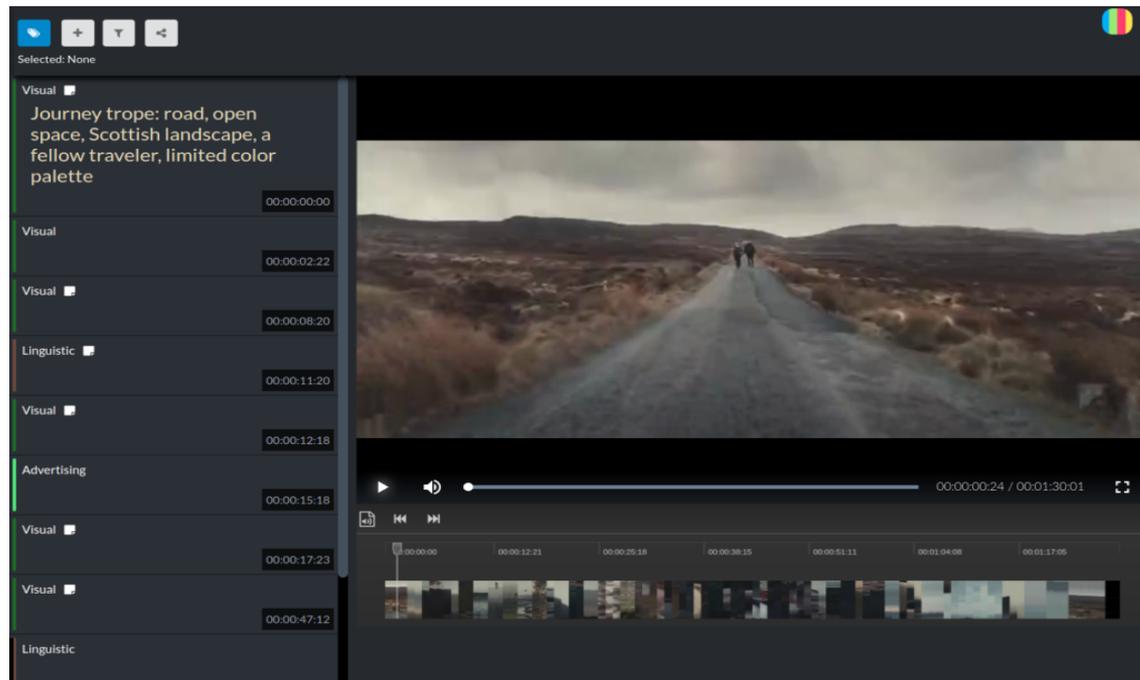


Using a Mind Map to draft a research question in Introductory Writing Courses



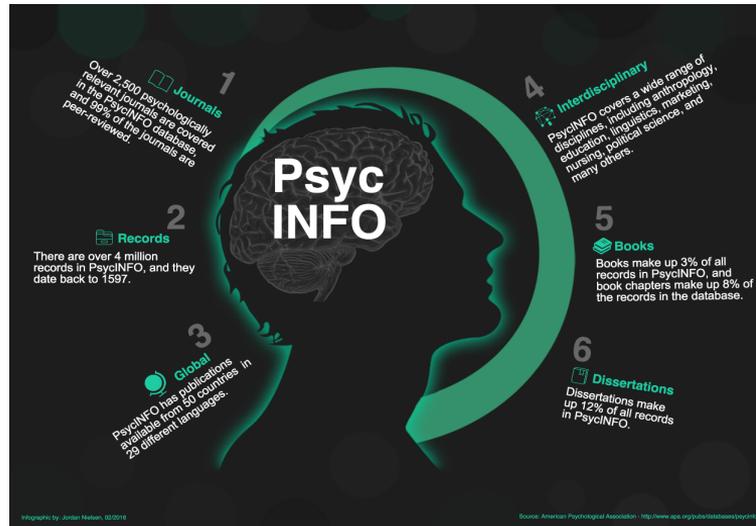
DIGITAL HUMANITIES

Mediate at
the
University
of
Rochester



MARKETING RESOURCES

Infographics to highlight library resources and market them to students



Learn a new language with Mango Languages!

11
Languages to
choose from.

French
Arabic (Iraqi)
Farsi (Persian)
American Sign Language
Dutch
Spanish (Latin America)
Korean
German
Japanese
Italian
Chinese (Mandarin)



Films that teach
language and culture.

Accents are sexy.



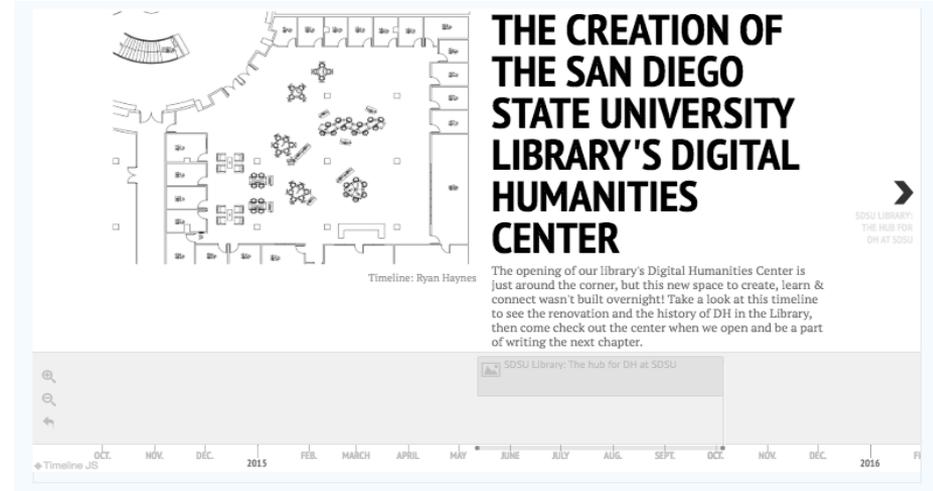
Perfect yours with Mango's
voice comparison feature.

Practice in your pajamas



Use Mango's free mobile app to study your favorite language from anywhere in the world—including at home in your pj's.

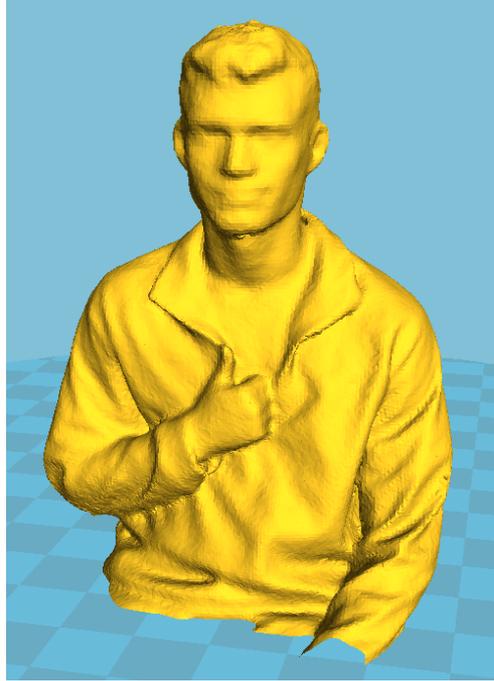
MARKETING SPACES



<https://library.sdsu.edu/digital-humanities-center>

3D SCANS

Using library resources to generate 3D scans to document physical objects

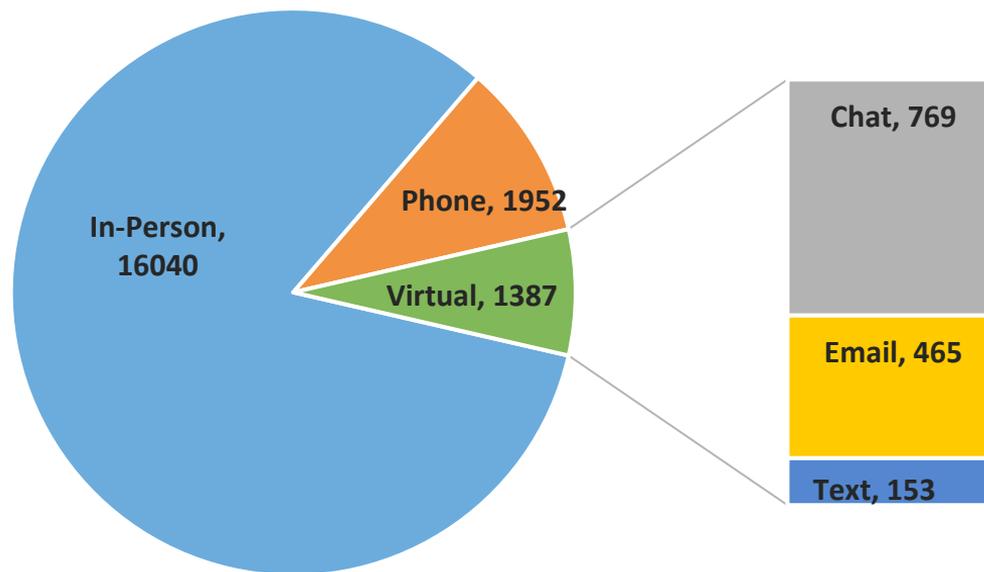


LIBRARY ASSESSMENT

REFERENCE SERVICE ASSESSMENT

Number of Reference Questions Asked

Usage
Statistics for
2014-2015
academic
year

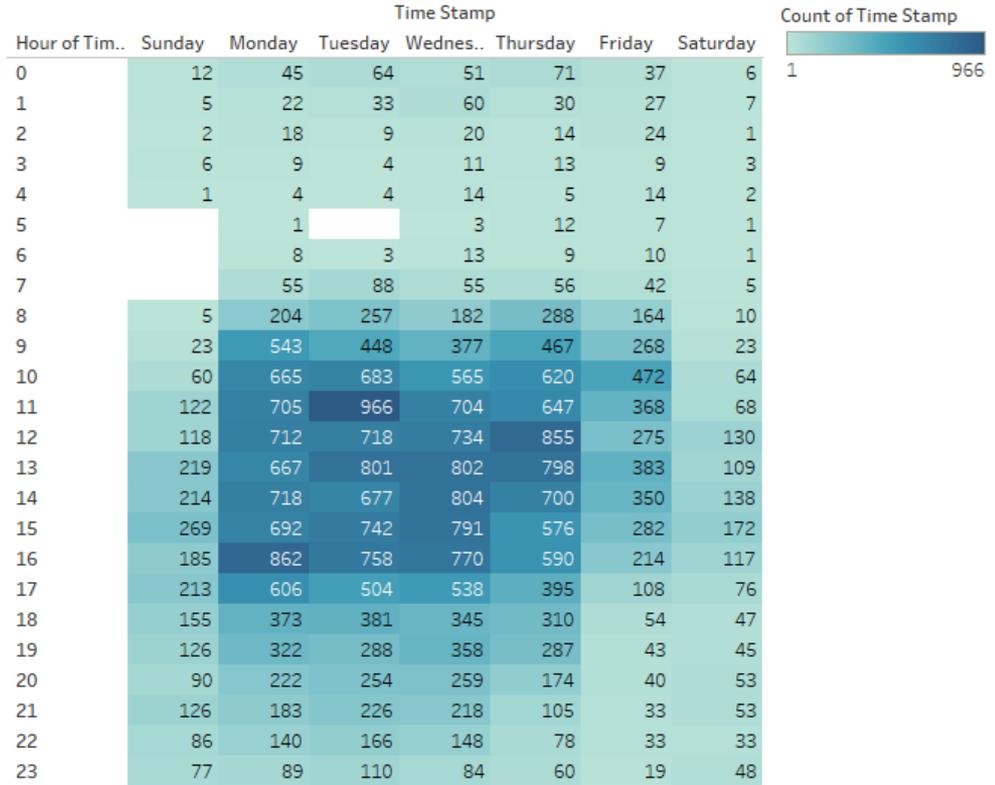


Categories of Questions asked at the Reference Desk



When People Search the Library's Homepage from On-Campus

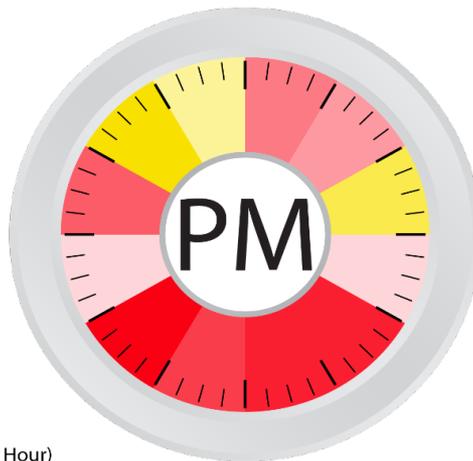
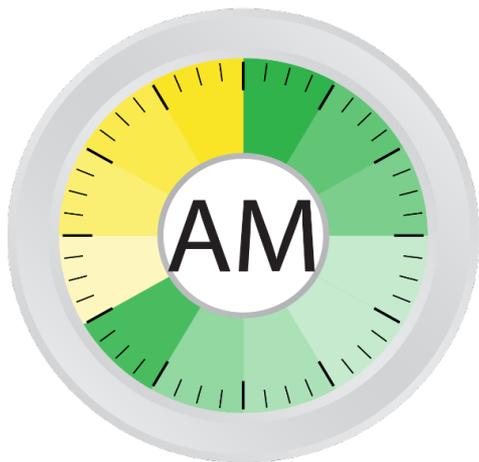
Sheet 2



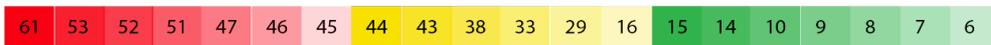
Count of Time Stamp broken down by Time Stamp Weekday vs. Time Stamp Hour. Color shows count of Time Stamp. The marks are labeled by count of Time Stamp.

CHAT SERVICE ASSESSMENT

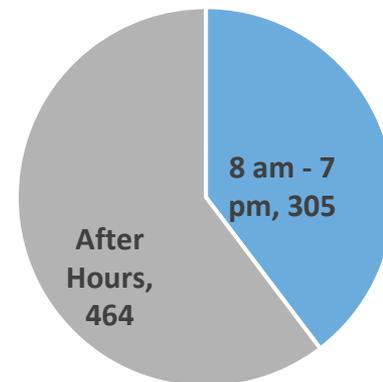
Number of Chat Questions Asked per Hour for the Academic Year, 2014 - 2015



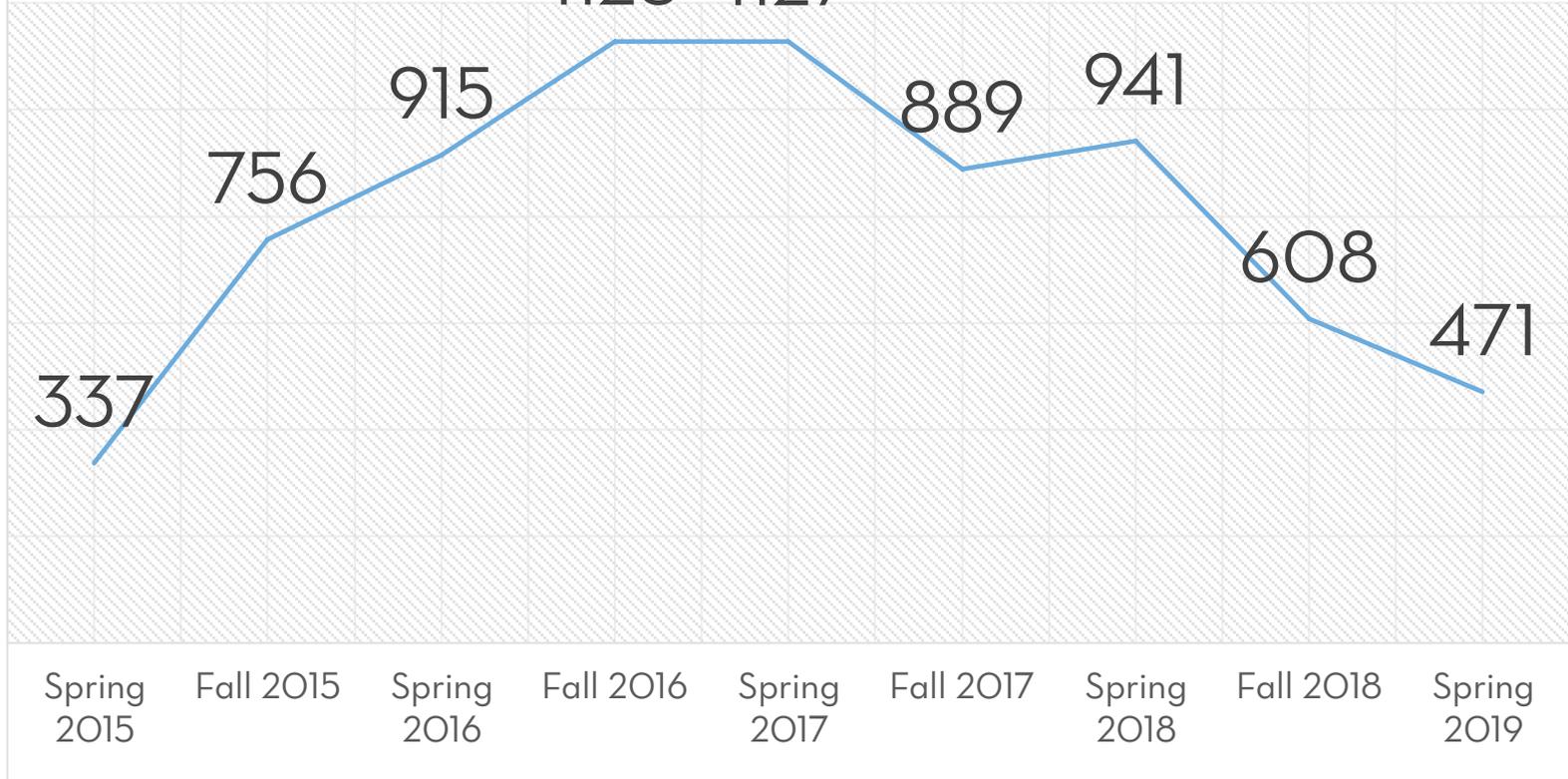
Scale (Number per Hour)



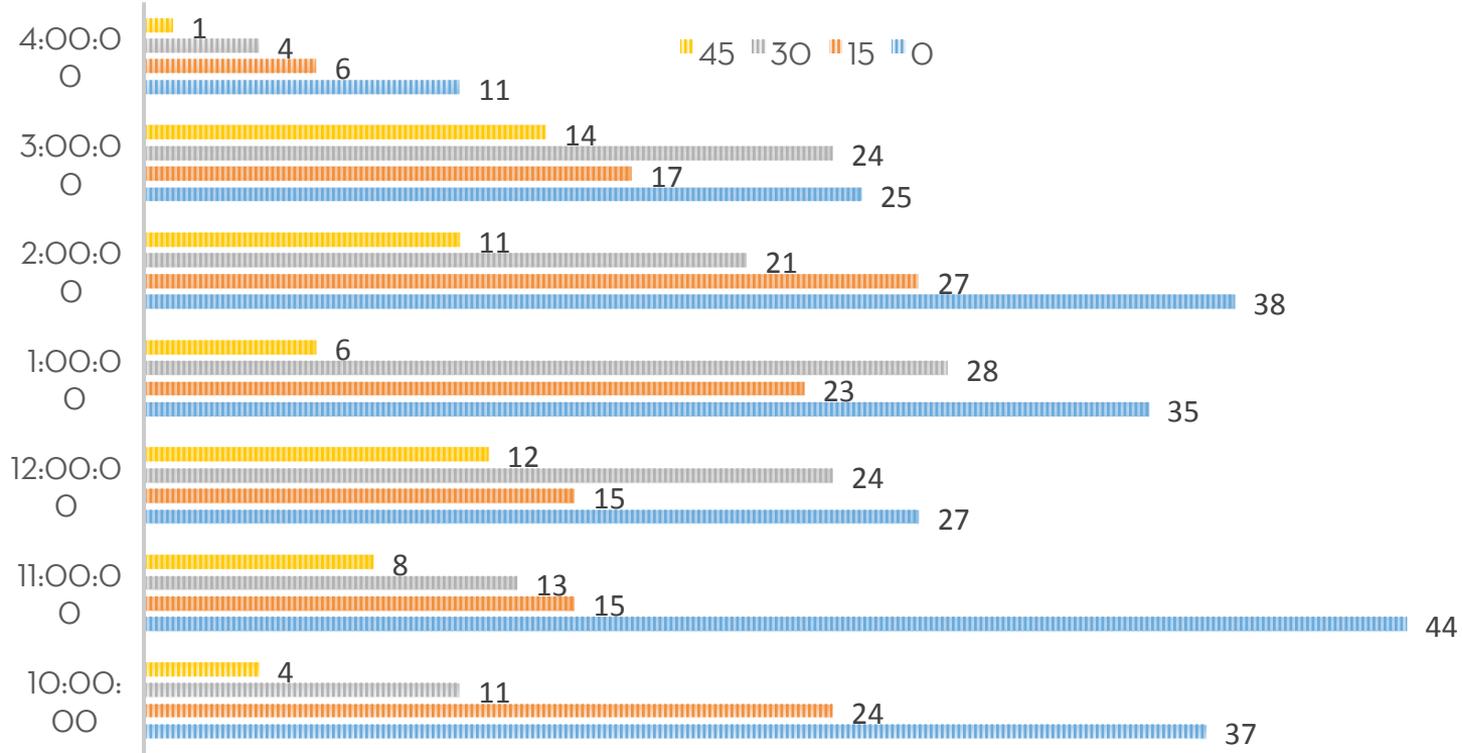
Chat Questions asked in relation to Reference Desk Hours



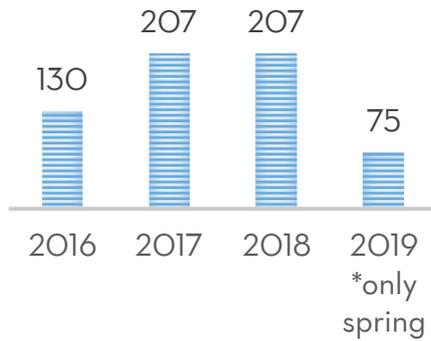
3D Print Requests Fulfilled



TIME OF DAY TRAININGS HAPPEN



ORIENTATION ATTENDANCE BY YEAR



ORIENTATION ATTENDANCE BY MONTH



LIBRARY RESEARCH

CASE STUDY OF OUTREACH EVENTS

HOW TO LEVERAGE YOUR ACADEMIC LIBRARY'S MAKERSPACE TO HELP ADVANCE YOUR LOCAL COMMUNITY

SAN DIEGO COUNTY, CA



Source: Geographic Research, Inc. (2017). Map with EASI Education Index 2016 Sales Potential Data. Retrieved June 12th, 2017, from SimplyMap database.

BACKGROUND

buildIT is the makerspace of the SDSU Library. While its purpose is to provide access to technology to students, faculty, and staff, it has helped create opportunities to promote STEM in the San Diego Region. As the librarian, I led the development of various outreach events. However, I make sure to involve SDSU students in every event. Their involvement can vary from co-developing instruction to manning the exhibit table. While the focus is on giving back to the community, the goal of these events is to increase the learning opportunities for SDSU students. Below is a sample from 2015 - 2017.

TEEN TECH WEEK: ROBOTICS CHALLENGE

📍 Poway Branch, San Diego County Library

📅 Mar 11, 2017, 2 hours

👤 Teens

- 1 Created an interactive display for teens to explore robotics through active play and learn about FIRST Robotics from local high schoolers. The mentors of the HS Robotics Club are SDSU.



1

JENNY WONG-WELCH



WHY

- Personal fulfillment of giving back to the community that I grew up in
- Supporting the University's Strategic Initiative of "Contributing to the Advancement of the San Diego Region"
- Increasing the networking opportunities and learning experiences for SDSU students
- Doing STEM education to under-represented communities
- Advocate for information literacy within STEM disciplines

ATEC SCIENCE SUMMER CAMP

📍 San Diego State University Library

📅 Aug 1 - Aug 12, 2016, 8 hours

👤 Camp Participant, Ages 7 - 14

- 1 Taught 3D printing, where the kids were able to keep a 3D printed. Some kids 3D modeled a part that was later 3D printed. SDSU students assisted with developing and teaching the curriculum.



3

EXPLORE SDSU

📍 San Diego State University Library

📅 Mar 14, 2015 & Mar 19, 2016, 8 hours

👤 Future Student, All Ages

- 1 Participated in the SDSU open house where local community members can learn about the campus. This event is marketed to potential students to explore and learn about SDSU.



4



MAKER FAIRE SAN DIEGO

📍 Balboa Park

📅 Oct 3-4, 2015 & Feb 1-2, 2016, 16 hours

👤 So. California Resident, All Ages

- 1 Tabbed during the San Diego-wide maker faire to promote the student work for the buildIT makerspace and network with other makers in the community.



5



TECH FORUM 3D SCANNING

📍 Tinkering Studio, Fleet Science Center

📅 Aug 13, 2016, 2 hours

👤 Museum Visitor, All Ages

- 1 Taught 3D printing and 3D scanning by demonstration of creating 3D scans of people and allowing them to keep the export file to print.



6



WE ARE STEM!

📍 The O'Farrell Charter School*

📅 Feb 25, 2016 & Apr 16, 2017, 2 hours

👤 Female Student, Ages 10-14

- 1 Middle School girls came to the buildIT makerspace to learn about emerging technology. SDSU students designed a 3D printed solar system bracelet that each girl got to paint and take home.



7

REWARD FIELD TRIP

📍 Boys & Girls Club of South County*

📅 Sep 16, 2016 & Apr 26, 2017, 2 hours

👤 Boys & Girls Club Kids, Ages 10 - 14

- 1 As a reward for completing an after-school tech program, kids came to the buildIT makerspace to explore and play with its vast collection of emerging technology.



8

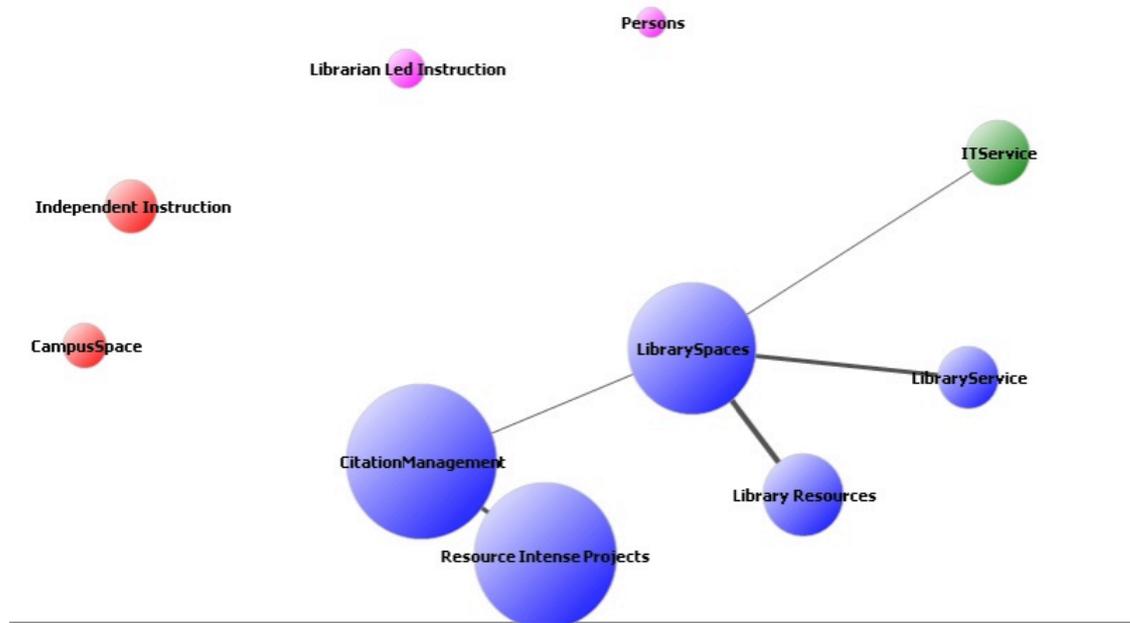


More Info:
<http://buildit.sdsu.edu/ala-2017/>

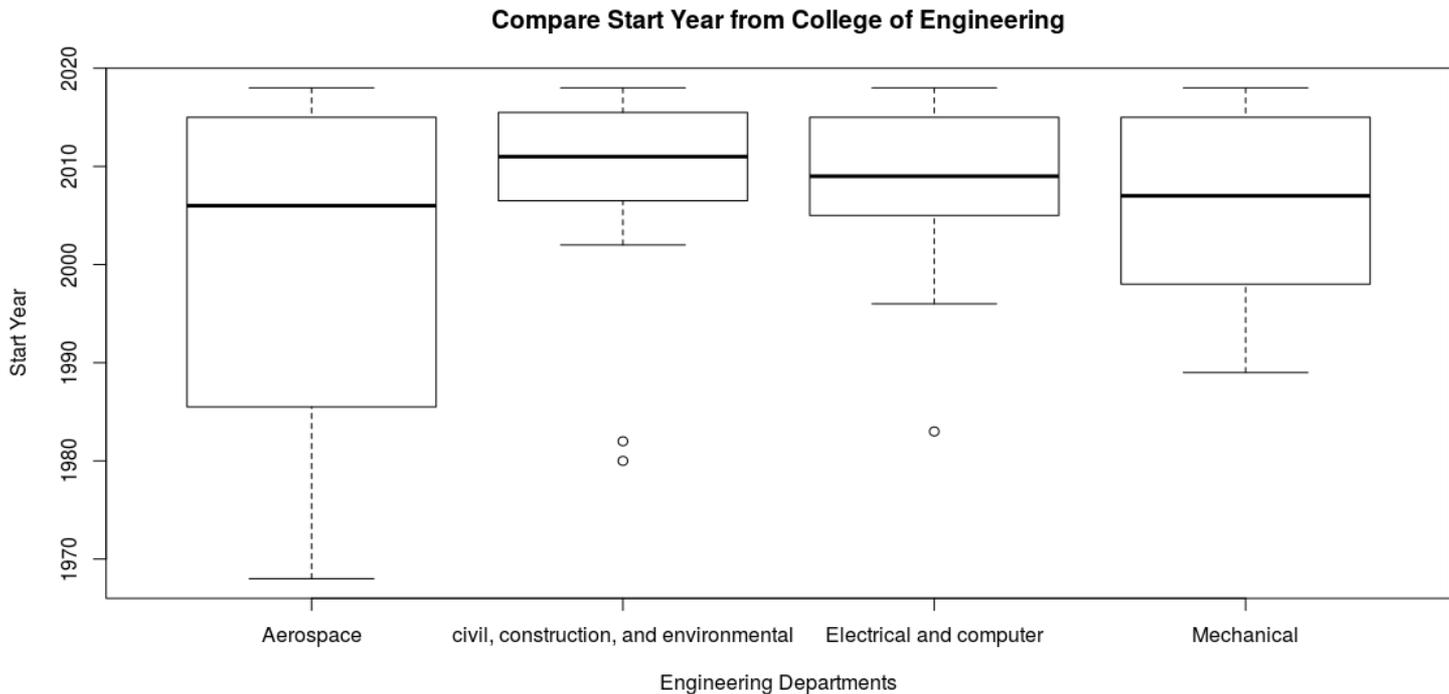
designed by freepik.com

SYLLABUS COLLECTION TEXT MINING ARTICLE

Category Occurrence and Relationships Map



CITATION ANALYSIS



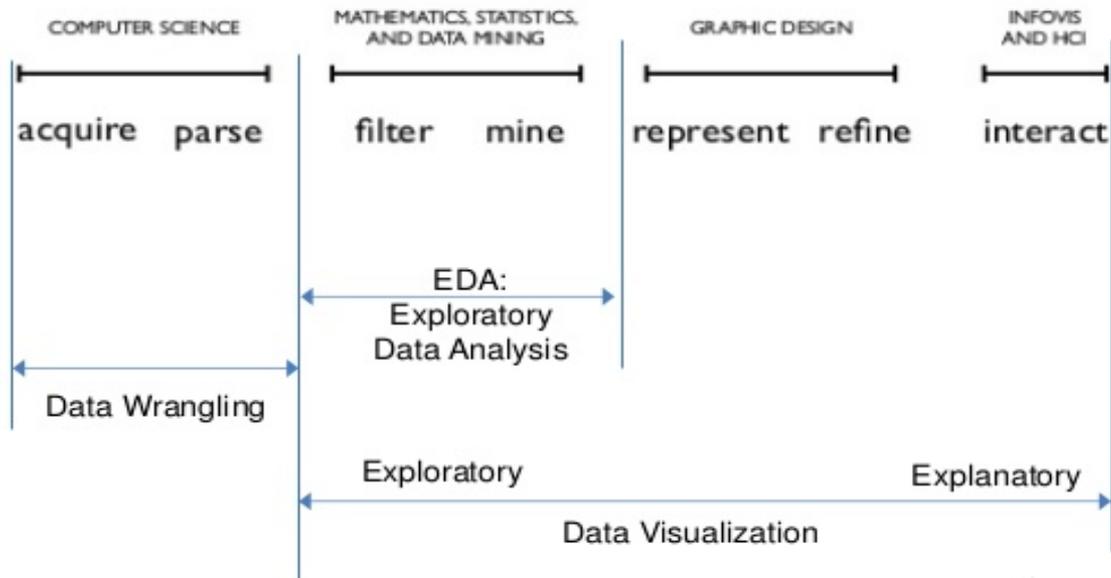
LITERATURE REVIEW

From a general search scan of “academic library” & “data visualization”

- Library Assessment
- Supporting library decision-making
- Examine an academic library collection
- Bibliometric/citation analysis
- Space management and usage – GIS & heat maps Seating patterns – heat map
- Data management
- Library data services – data management & teaching how to make data visualizations

DATA SCIENCE PROCESS

BEN FRY'S PROCESS



Source: *Computational Information Design* | Ben Fry



DATA ACQUISITION

WHAT DATA IS AVAILABLE?

- Usage
 - Gate Counts
 - Computer Log-ins
 - LibGuides
 - Primo Searches
 - Service Desks
 - Research Consultations
 - Instruction Sessions
 - Alma Circulation Stats
 - Special Collections Access
 - Journal COUNTER
- Collections
 - Alma Catalog
 - Vendor Supplied Info
- Administration
 - Budget
 - Staffing
- Campus
 - Student/Faculty Demographics
 - Research Output
- “Open” Sources
 - Data.gov, wikidata.org, datahub.io
 - Simply Analytics

Other Available Data?

LIBRARY DATABASES RESOURCES

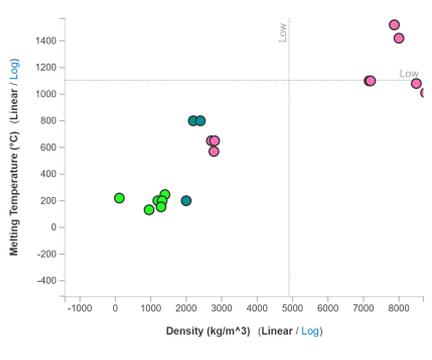
<https://libguides.sdsu.edu/az.php?t=21125>

Select Materials **23 selected**

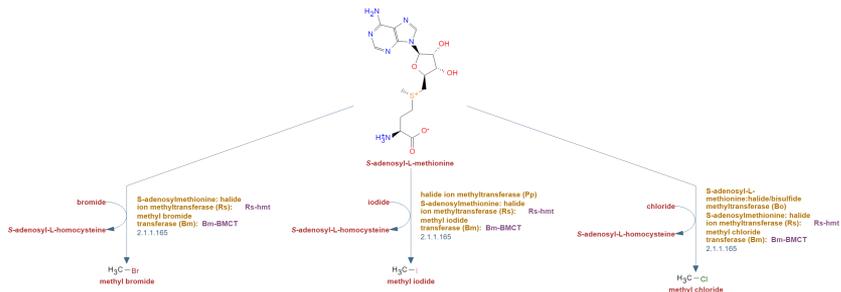
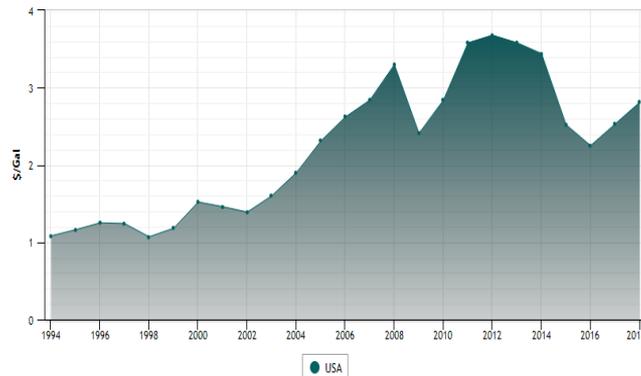
Enter a material or classification

Expand All Deselect All

- ▶ Metal (10)
- ▶ Polymer (6)
- ▶ Ceramic (3)
- ▶ Composite (4)



Gasoline Prices - All Grades from the Retail Gasoline Prices Database
USA - 1990 - 2018
Data Planet™. A SAGE Publishing Resource, Source: Energy Information Administration
Unit: \$/Gal



COLLECTING YOUR OWN DATA

- Defining your own data collection procedures is incredibly powerful
- Collecting one's own data lets data processing inform collection



Source: Magnuson, L. (2016). Data visualization: A guide to visual storytelling for libraries (LITA guides). Lanham: Rowman & Littlefield

EXPLORATORY DATA ANALYSIS

Parsing, Filtering, Mining

EXAMINING THE DATA

- Do you have all the data you need? Does it include all the variables that you are interested in?
- Are there any obvious errors in your data? Is there any data that is missing?



UNDERSTANDING THE DATA TYPES

- What type of data have you acquired?

Types	Examples
Categorical nominal	Countries, gender, text
Categorical ordinal	Olympic medals, "Likert" scale
Quantitative (interval-scale)	Dates, temperature
Quantitative (ratio-scale)	Prices, age, distance

- What is the range of values for each type of data?

Data	Types	Range
Event	Quantitative (interval-scale)	27 different years (1896–2012)
Medal	Categorical ordinal	Gold, silver, bronze
Athlete	Categorical nominal	1500+ different athlete names
Result	Quantitative (ratio-scale)	Race results (9.59s > 4:02:59)
Country	Categorical nominal	96 different country names

PARSING & FILTERING

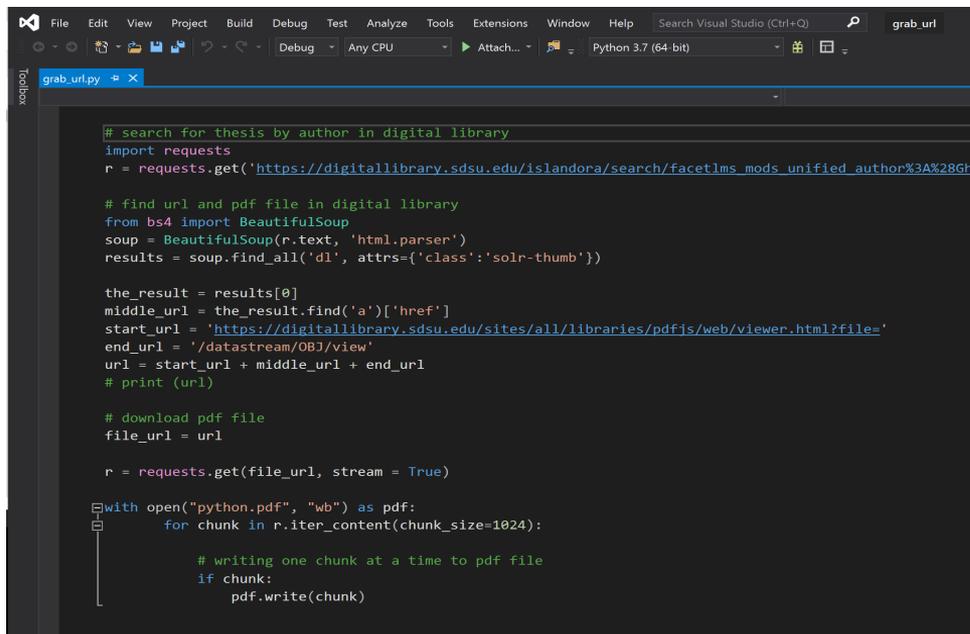
- Data sets may need to be supplemented, pruned, or aggregated.
- Always keep in mind that we refine our data sets to *enhance* their objectivity, not discard it
- Discarding data points simply because they are problematic is unethical and distorted
- Example:
 - Research a library's virtual reference service
 - Filter data outside the scope such as interactions via phone or in person
- Documentation is vital
 - Should maintain a running narrative of the alterations we make and their rationale

TRANSFORMING DATA

- Transforming for Quality
 - Do you need to clean up your data? Do you need to fix any errors or fill in any gaps in your data?
- Transforming for Analysis
 - Parsing (splitting up) and variables, such as extracting year from a date value
 - Merging variables to form new ones, such as creating a whole name out of title, forename, and surname
 - Converting qualitative data/free-text into coded values or keywords
 - Deriving new values out of others, such as gender from title or a sentiment out of some qualitative data
 - Creating calculations for use in analysis, such as percentage proportions
 - Removing redundant data for which you have no planned use (be careful though!)

PARSING/FILTERING TOOLS

- Excel
- R
- OpenRefine
- Scripting Language
 - Python
 - Perl



```
# search for thesis by author in digital library
import requests
r = requests.get('https://digitallibrary.sdsu.edu/islandora/search/facetlms_mods_unified_author%3A%28Gh

# find url and pdf file in digital library
from bs4 import BeautifulSoup
soup = BeautifulSoup(r.text, 'html.parser')
results = soup.find_all('dl', attrs={'class': 'solr-thumb'})

the_result = results[0]
middle_url = the_result.find('a')['href']
start_url = 'https://digitallibrary.sdsu.edu/sites/all/libraries/pdfjs/web/viewer.html?file='
end_url = '/datastream/OBJ/view'
url = start_url + middle_url + end_url
# print (url)

# download pdf file
file_url = url

r = requests.get(file_url, stream = True)

with open("python.pdf", "wb") as pdf:
    for chunk in r.iter_content(chunk_size=1024):
        # writing one chunk at a time to pdf file
        if chunk:
            pdf.write(chunk)
```

BIBLIOMINING

- Or data mining for libraries, was first used by Nicholson & Stanton (2003) to describe the combination of data warehousing, data mining and bibliometrics. This term is used to track patterns, behavior changes, and trends of library systems transactions.

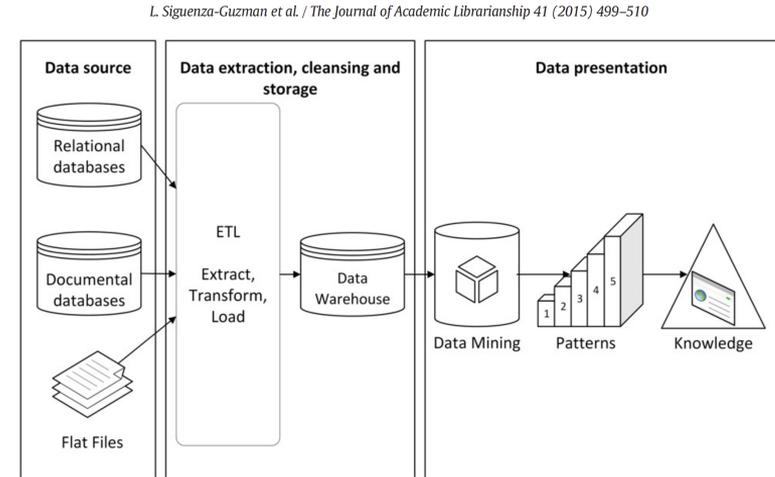


Fig. 1. Data mining process, based on Han et al. (2011).

Source: Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. *The Journal of Academic Librarianship*, 41(4), 499–510. <https://doi.org/10.1016/J.ACALIB.2015.06.007>

HOLISTIC APPROACH TO LIBRARY EVALUATION

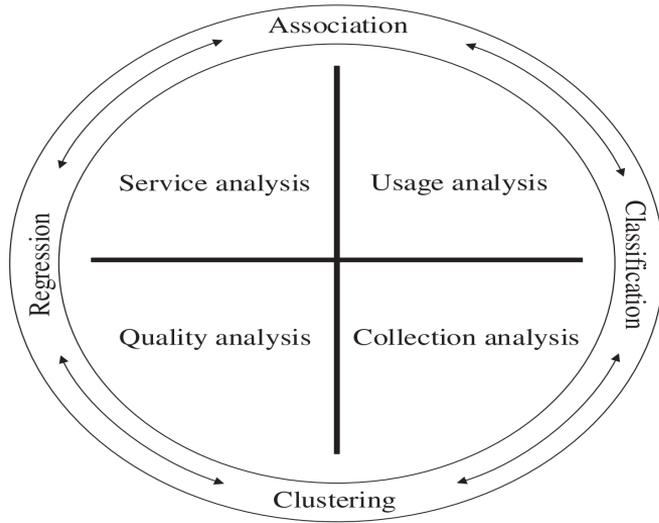


Fig. 2. Classification framework for data mining techniques based on the Ngai et al. (2009) approach.

1. Internal perspective of the library system – process/ service analysis
2. External perspective of the library system – quality analysis
3. Internal perspective of the library collection – collection analysis
4. External perspective of the library collection – usage analysis

Source: Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. *The Journal of Academic Librarianship*, 41(4), 499-510. <https://doi.org/10.1016/J.ACALIB.2015.06.007>

DATA MINING MODELS

1. Association

- Aims to find the existing (or potential) relationships between data items –
- statistics and apriori algorithms

2. Clustering

- task of uncovering unanticipated trends by segmenting no predefined clusters
- neural networks, k-means algorithms, and discrimination analysis

3. Classification

- the task of attempting to discover predictive patterns by classifying database records into a number of predefined categorical classes based on certain criteria
- neural networks, decision trees, and if-then-else rules

4. Regression

- statistical technique that maps a data item to a real-valued prediction variable
- used to capture the trends of frequent patterns
- linear regression and logistic regression analysis

FORMAT/PROCESS



- The technical procedures of converting data from their original format to the one that our visualization software utilizes
- As we go along converting data from one format to another, it behooves us to keep intermediary copies of the data as they transition.
 - Git

A screenshot of a GitHub repository page for 'SDSUbuildIt / request-scripts'. The page shows the repository name, navigation tabs (Code, Issues, Pull requests, Projects, Wiki, Security, Insights, Settings), and repository statistics (7 commits, 2 branches, 0 releases, 0 contributors). A commit history table is visible, listing commits by Lindsay White. The current commit selected is 'request-scripts', which contains a README.md file. The README content is displayed below, describing the scripts for the request form for 3D printing.

SDSUbuildIt / request-scripts

Unwatch 2 Star 1 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

The scripts for the request form for 3D printing. Edit

Manage topics

7 commits 2 branches 0 releases 0 contributors

Branch: master New pull request Create new file Upload files Find File Clone or download -

Commit	Message	Time
Lindsay White	Update README.md	Latest commit bb45831 on Sep 10, 2015
Lindsay White	ColorName	4 years ago
Lindsay White	Create Emails	4 years ago
Lindsay White	HideRows	4 years ago
Lindsay White	Update README.md	4 years ago
Lindsay White	UpdatePrintTotals	4 years ago
Lindsay White	Create changeDateCompleted	4 years ago

README.md

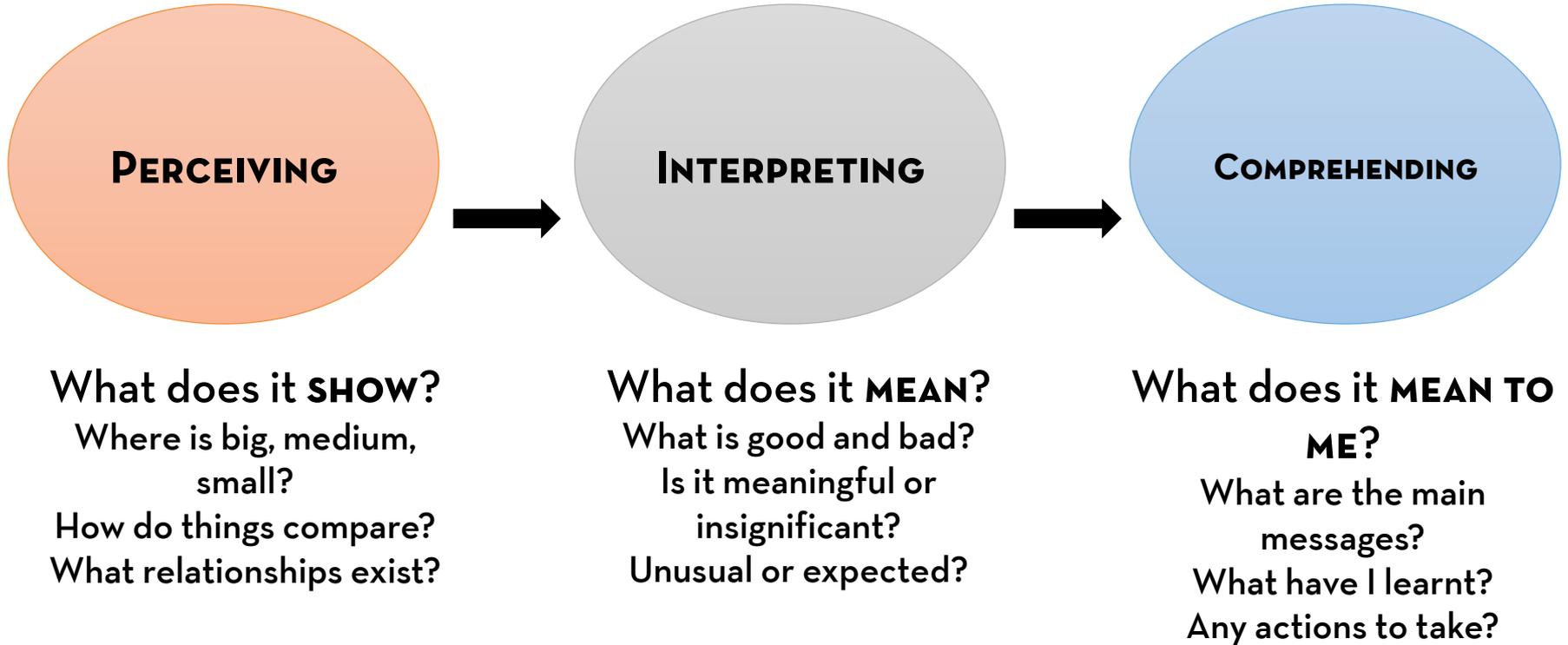
request-scripts

The scripts for the request form for 3D printing.

These are written in Google's flavor of javascript for Google Docs. They are used with the Google Sheets spreadsheet that is linked to the form that SDSU students, faculty, and staff can use to request 3D prints.

MAKING A DATA VISUALIZATION

THE THREE STAGES OF UNDERSTANDING



THE THREE PRINCIPLES OF GOOD VISUALIZATION DESIGN

PRINCIPLE 1

Good data visualization is
TRUSTWORTHY

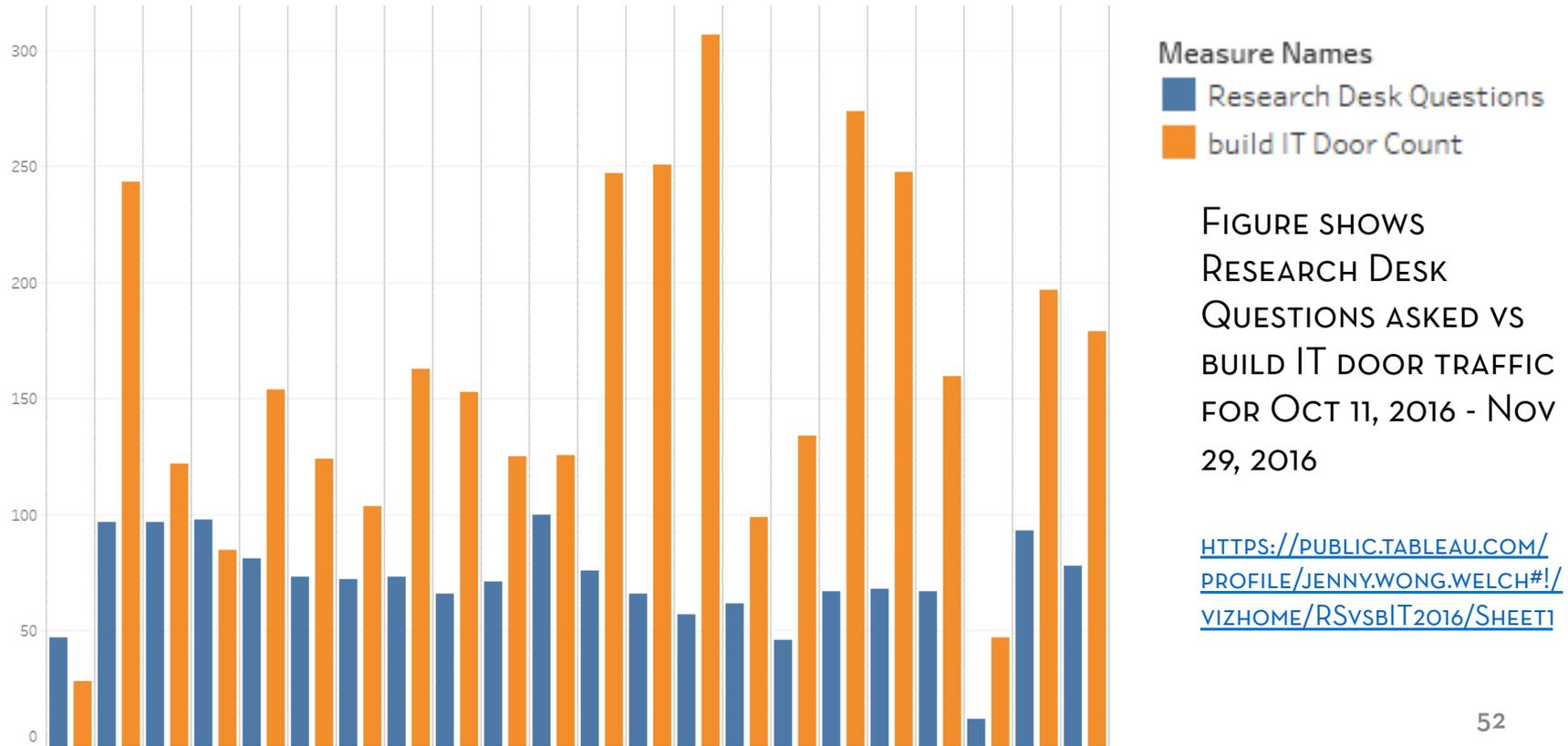
PRINCIPLE 2

Good data visualization
is
ACCESSIBLE

PRINCIPLE 3

Good data visualization
is
ELEGANT

LYING WITH DATA



Measure Names

- Research Desk Questions
- build IT Door Count

FIGURE SHOWS
RESEARCH DESK
QUESTIONS ASKED VS
BUILD IT DOOR TRAFFIC
FOR OCT 11, 2016 - NOV
29, 2016

[https://public.tableau.com/
profile/Jenny.Wong.Welch#!/
vizhome/RSvsBIT2016/Sheet1](https://public.tableau.com/profile/Jenny.Wong.Welch#!/vizhome/RSvsBIT2016/Sheet1)

DATA STORIES

The simplest story – counting or totaling something

- Measurement

But it's often difficult to know if that's a lot or a little. For that, you need context – which can be provided by:

- Proportion
- Internal comparison
- External comparison
- Change over time
- 'League tables'
- Analysis by categories

Or you can relate factors numerically

- Association

DATA STORIES CONT.

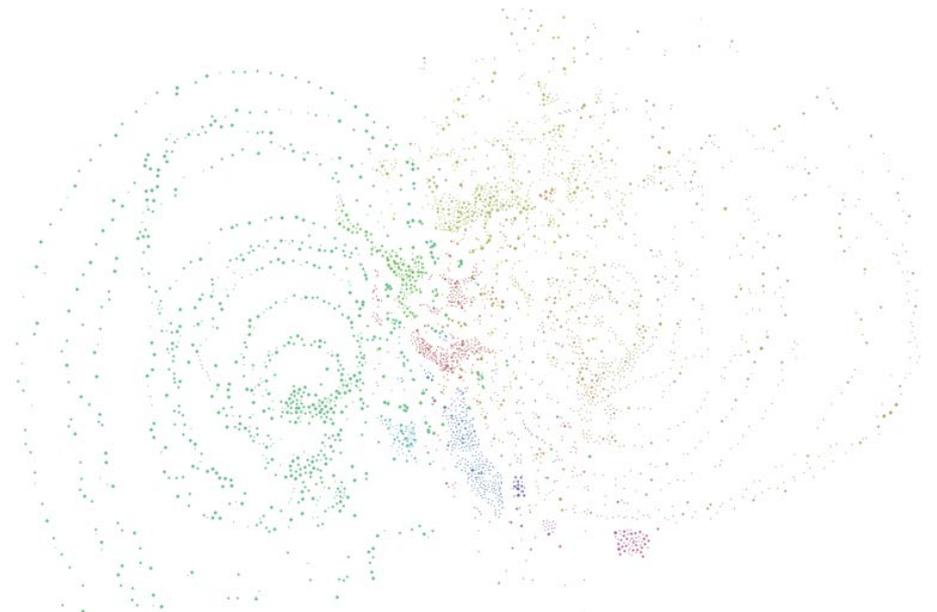
But, of course, always remember that correlation and causation are not the same thing.

So if you're investigating paper clip spending, are you also getting the following figures:

- Total spending to provide context?
- Geographical/historical/other breakdowns to provide comparative data?
- The additional data you need to ensure comparisons are fair, such as population size?
- Other data which might provide interesting analysis to compare or relate the spending to?"

STORYTELLING WITH DATA: A DATA VISUALIZATION GUIDE FOR BUSINESS PROFESSIONALS

- Understand the context
- Choose an appropriate visual display
- Eliminate clutter
- Focus attention where you want it
- Think like a designer
- Tell a story



TYPES

- **1D/Linear**
 - lists of data items, organized by a single feature (e.g., alphabetical order)
(not commonly visualized)
- **2D/Planar** (incl. Geospatial)
- **3D/Volumetric**
 - 3D computer models, surface and volume rendering, computer simulations
- **Temporal**
 - Timeline, time series, connected scatter plot, Gantt Chart,
- **nD/Multidimensional**
 - Category proportions, counts
 - Pie chart, histogram, wordle/tag cloud, unordered bubble chart/cloud, bar chart, tree map, scatter plot, line chart, step chart, heat map, radar/spider chart, box and whisker plot,
- **Tree/Hierarchical**
 - General tree visualization, dendrogram, radial tree, hyperbolic tree, tree map, wedge stack graph (radial hierarchy)/sunburst
- **Network**
 - Matrix, node-link diagram, dependency graph/circular hierarchy, hive plot, subway map

https://guides.library.duke.edu/datavis/vis_types &

This taxonomy is based on a data taxonomy from: Shneiderman, B. (1996).

[The eyes have it: A task by data type taxonomy for information visualizations.](#) *Proceedings of IEEE Symposium on Visual Languages - Boulder, CO* (pp. 336-343).

DATA VISUALIZATION TOOLS

- Simply Analytics
- Tableau
- Raw.DensityDesign.org
- Bubblus
- The Noun Project
- Freepik
- Piktochart
- Excel
- Adobe Illustrator
- R
- D3

MAKE A DATA VISUALIZATION IN A HOT MINUTE

<http://raw.densitydesign.org/>

<https://tinyurl.com/csudataviz>

- 3D Printer Requests & Usage for Spring 2018
- Data Filtering
 - Deleted Blanks for timestamp, when printed
 - Delete print duration of 0:00:00
 - Undefined for blanks in why used
 - Split timestamp to date and time columns
 - Added columns - submission day & days between submission and print

- Data Dashboards
 - Alma
 - Primo
 - LibInsights
 - <https://sdsu.libinsight.com/dataseta.php?id=16539>

HTTPS://TINYURL.COM/
CSUDATAVIZ

LET'S CREATE A DATA VISUALIZATION

- Topic?
- Audience?
- Communication Objective?
- What's the story in my data?
- Why am I using a data visualization?
- Colors, icons, images mean symbolic meaning?
- Type of Graph?

DATA VISUALIZATION WORKSHEET	
CSU Lib IT 2019, Northridge, CA	
You will create a data visualization based on your own research findings or data you have identified about your topic. Your goal is to increase understanding of your data and use the data to tell a story, while applying principles of good data visualization. You may want to use this worksheet to begin planning your data viz.	
Topic:	
Audience:	
Communication Objective:	
What's the story in my data?	
Why am I using a data visualization to explain this story?	
What colors (if any) have symbolic meaning—both positive and negative—for this topic?	
What are some icons or images that have symbolic meaning for this topic?	
What type of graph can convey this data?	

PUBLISHING

- If we want our research study to be verifiable, we *must* publish the data in some form alongside our final visualization.
 - Means that others can correct errors in the event that our filtering/refining was imperfect, remix our data by applying it to another visualization technique, or reuse it as part of a larger study that takes our data as but one of several sources
- Considerations:
 - Wary of publishing any personally identifiable information (PII) or data with strong patterns that could be linked to individuals
 - Patron privacy - should not share or publish data that could reveal information about an individual without their explicit permission
 - 73 questions in July vs question at 7:05 pm on July 3 on illegal fireworks
- Document thoroughly → a raw data set with no description of its schema, data types, or collection procedures is often worse than no data at all because of the ease with which it is misinterpreted

BIBLIOGRAPHY

- Kirk, A. (2016). Data visualisation: A handbook for data driven design.
- Magnuson, L. (2016). Data visualization: A guide to visual storytelling for libraries (LITA guides). Lanham: Rowman & Littlefield.
- Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. *The Journal of Academic Librarianship*, 41(4), 499-510. <https://doi.org/10.1016/J.ACALIB.2015.06.007>
- <https://guides.library.duke.edu/datavis>