Teaching Data: Developing Data Instruction Using a Multi-Level Competency Model

Laying the Foundation for Research Data Services: Session 2

Texas Digital Library

PREVIOUSLY ON...
Laying the Foundation for Research Data Services

SESSION 1 – Getting Started with the Texas Data Repository and Data Competencies

- ◦ This initial webinar will cover the Texas Data Repository's role as an integral part of the research data infrastructure in Texas and introduce a set of competencies for data training on your campus. This session will preview following sessions and include results from the data symposium and how these webinars address some of these needs.
- ◦ Recording: http://hdl.handle.net/2249.1/79231

Texas Digital Library

Teaching Data

Developing Data Instruction Using a Multi-Level Competency Model

Peace Ossom Williamson
Director, Research Data Services
University of Texas at Arlington Libraries

What?

Data Information Literacy
datainfolit.org

- IMLS-funded research project (Purdue, Minnesota, Oregon, Cornell)

- Aim: develop and implement curriculum around data information literacy

Why?

Campus data needs

- Information/data ethics

- Integration of data in assignments & data literacy

- Improved research methods

- Digital literacy

Data Are more than numbers

Examples:
- Audio recordings
- Images & photographs
- Measurements
- Addresses & geocoordinates
- Text / Corpora

Data Literacy

the ability to read, create, utilize, communicate, and criticize data.

Phase 1
Development

Adapting DIL competencies to meet the needs of students, faculty, and staff of UTA campus in regards to working with data for teaching, learning, and research.

Phase 1:
12 Initial Competencies

1. Intro to Datasets & Data Formats
2. Data Discovery & Acquisition
3. Data Management & Organization
4. Data Conversion & Interoperability
5. Quality Assurance
6. Metadata
7. Data Curation & Re-Use
8. Cultures of Practice
9. Data Preservation
10. Data Analysis
11. Data Visualization
12. Ethics, including citation of data

http://blogs.lib.purdue.edu/dil/the-twelve-dil-competencies

Phase 1:
Data Literacy Competencies Development

[Data information literacy competencies](#) were adjusted to include

- Levels: emerging, intermediate & expert

- Additional, removed & re-defined competencies

- Learning outcomes for each

Phase 1:
DIL Competencies: QA, Metadata

- ## Metadata:
    - Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others.
    - Develops the ability to read and interpret metadata from external disciplinary sources.
    - Understands the structure and purpose of ontologies in facilitating better sharing of data.
- ## Data description: Able to read & interpret metadata, understands structure & purpose of ontologies, etc.

Phase 1:
DIL Competencies: QA, Metadata

- **<u>Quality Assurance:</u>** Recognizes and resolves any apparent artifacts, incompletion, or corruption of data sets. Utilizes metadata to facilitate understanding of potential problems with data sets.
- <u>Data quality & documentation:</u>
  - Documents data sufficiently to enable reproduction of research results & data by others
  - Recognizes, documents, and resolves any apparent, artifacts, incompletion, or corruption of data

Phase 1:
DIL Competencies: Cultures of Practice

- Recognizes practices, values, and norms.

- Recognizes relevant data standards of field.

Phase 1:
DIL Competencies: Cultures of Practice

Instead...

- <u>**Databases & data formats:**</u> Becomes familiar w/ standard data formats and types for the discipline

- <u>**Discovery & acquisition of data:**</u> Locates and utilizes disciplinary data sources and repositories

- <u>**Data structuring & cleaning:**</u> Understands practices, values, and norms of (sub)discipline as they relate to structuring data

Phase 1:
Competency Example

Ethics		
Emerging	**Intermediate**	**Expert**
- Develops understanding of attribution and reuse - Practices citing data - Develops mindfulness in data ownership and privacy	- Understands the ethos of the discipline when it comes to sharing and administration of data - Practices citing data - Explains attribution and reuse	- Recognizes the role of IP, privacy, and confidentiality issues - Creates data sharing agreements for data that cannot be shared w/o stipulations - Applies best practices for anonymizing data & reducing risk of re-identification of personal info - Makes own data citable

Phase 1:
Competency Example

Data Description		
Emerging	Intermediate	Expert
• Understands the rationale for descriptive metadata	• Employs basic descriptive, structural, and administrative metadata • Knows how to capture basic metadata elements • Develops understanding of when and where to deploy metadata	• Develops structures for customized descriptive, structural, and administrative metadata • Analyzes and interprets metadata from external disciplinary sources • Understands the structure and purpose of ontologies and metadata interoperability in facilitating better data sharing

Phase 1:
Competency Example

Data Sharing and Preservation	
Emerging	Intermediate
Recognizes the benefits of data preservation - Recognizes the practices, values, and norms of (sub)discipline as they relate to sharing & preserving data	- Is able to distinguish between active data and stored data - Understands basic definitions & processes in data preservation **Recognizes benefits & costs of preservation** - Develops understanding of which elements of a dataset are likely to have future value for self and others - Is able to determine when and how to backup data

Phase 1:
Competency Example

Data Sharing and Preservation
Expert
- Evaluates the benefits and costs of data preservation   - Identifies potential opportunities and improvements of data curation and sharing practices in (sub)discipline   - Utilizes best practices in preparing data for its eventual preservation throughout its active lifecycle   - Articulates the potential long-term value of own data for self or for others and is able to determine an appropriate preservation timeframe   - Discusses the cost and complexity of data curation as well as its vitality to community-driven e-research and to reproducibility

Phase 1:

Draft - Data Literacy Competencies 2.0

1. Data Awareness & Knowledge
2. Discovery & Acquisition of Data
3. Databases & Data Formats
4. Data Conversion & Interoperability
5. Data Organization & Management
6. Data Wrangling
7. Data Processing & Analysis
8. Data Quality & Documentation
9. Data Description
10. Ethics
11. Data Visualization & Representation
12. Data Sharing & Preservation

Phase 1:
Draft - Data Literacy Competencies 2.0

1. Data Awareness & Knowledge
2. Discovery & Acquisition of Data
3. Databases & Data Formats
4. Data Conversion & Interoperability
5. Data Organization & Management
6. Data Wrangling

7. Data Processing & Analysis
8. Data Quality & Documentation
9. Data Description
10. Ethics
11. Data Visualization & Representation
12. Data Sharing & Preservation

Phase 2

Implementation

Data Literacy Sessions are taught at workshops and in a number of courses.

Phase 2:

Support Services Implementation

Instruction scaffolding in classes and workshop partnerships with graduate & faculty training groups are being coordinated for campus.

Courses:

- English
- Public Health
- Journalism
- Community Health Nursing

Data Organization & Management

the development and execution of architectures, policies, practices, and procedures that properly manage the full data lifecycle needs of a project.

WHY CARE?

IS THERE A REPRODUCIBILITY CRISIS?

7% Don't know
52% Yes, a significant crisis
3% No, there is no crisis
1,576 researchers surveyed
38% Yes, a slight crisis
©nature

Retrieved from http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970
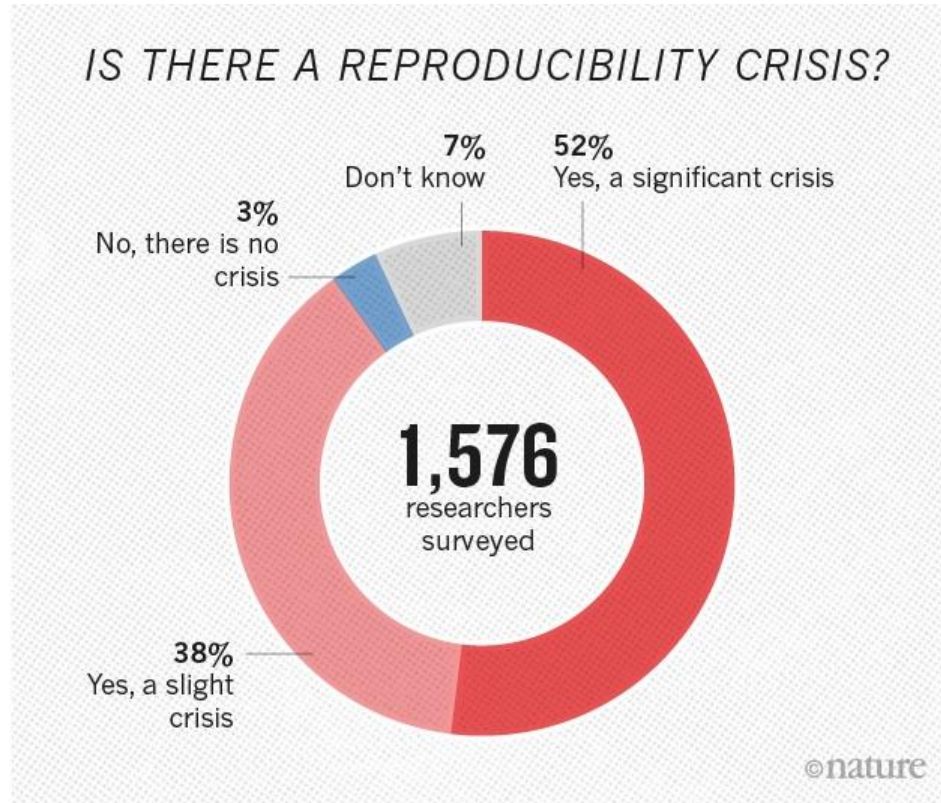
Ethics

Ethics – the practices centering on data ownership, confidentiality, and privacy
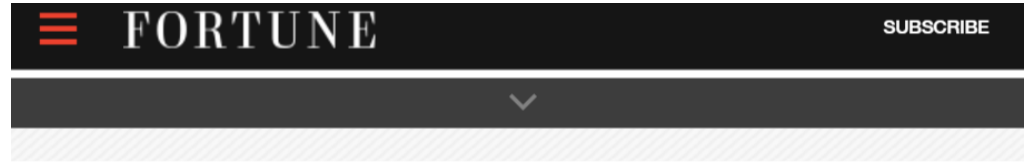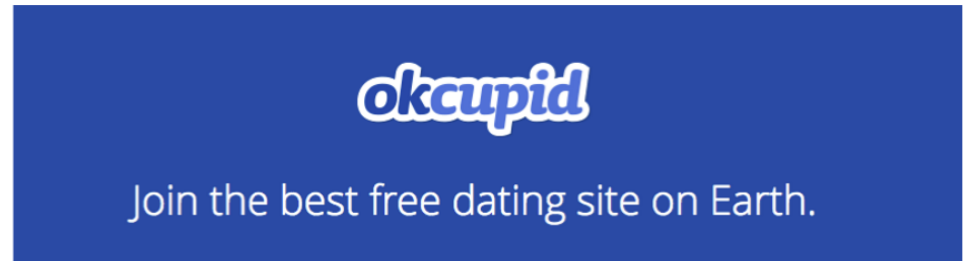
"We thought this was an obvious case of public data scraping so that it would not be a legal problem,"

Kirkegaard wrote to *Fortune*.

confidentiality - protection of information about a person

privacy - protection of the person

Questions to ask

1. How do I ensure the right to consent for individuals and communities?

2. How do I preserve privacy, security, and ownership around their data?

Tech Moves Faster Than Laws

Data Visualization & Representation

The development of charts, maps, infographics, and other visual displays of data using best practices and meaningful use of visualization programs.

Narrative   Structure & Flow   Visual Communication

Data Visualization & Representation

The development of charts, maps, infographics, and other visual displays of data using best practices and meaningful use of visualization programs.

Gun deaths in Florida

Number of murders committed using firearms

2005
Florida enacted its 'Stand Your Ground' law

873

721

1990s   2000s   2010s
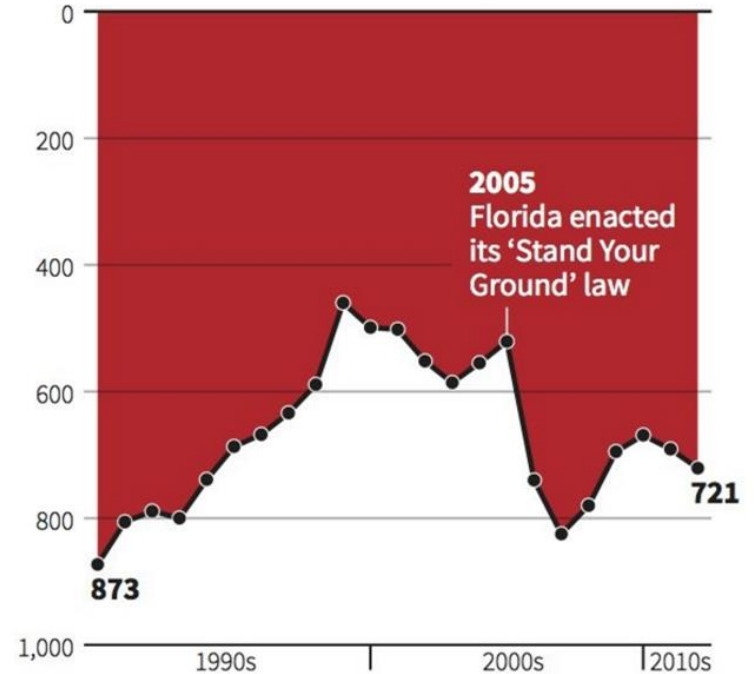
Source: Florida Department of Law Enforcement

http://www.visualisingdata.com/2014/04/the-fine-line-between-confusion-and-deception

IF YOU TORTURE DATA LONG ENOUGH ,
IT WILL CONFESS TO ANYTHING .

What's wrong?

Interest Rates

Year	Rate
2008	3.141%
2009	3.142%
2010	3.144%
2011	3.147%
2012	3.152%
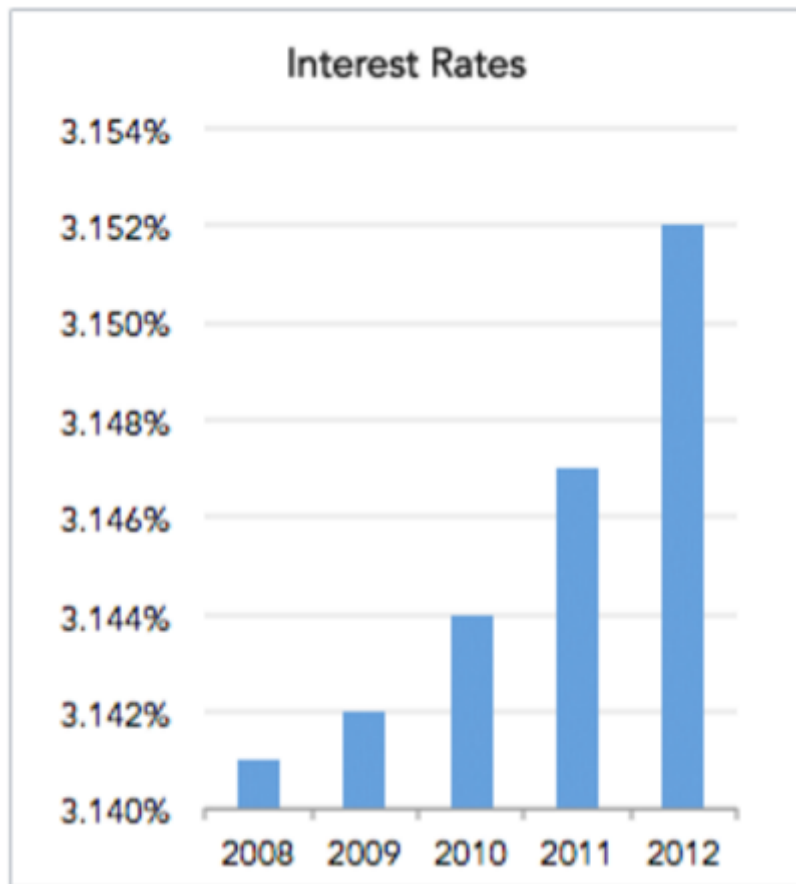
Less dramatic than you thought?

Interest Rates
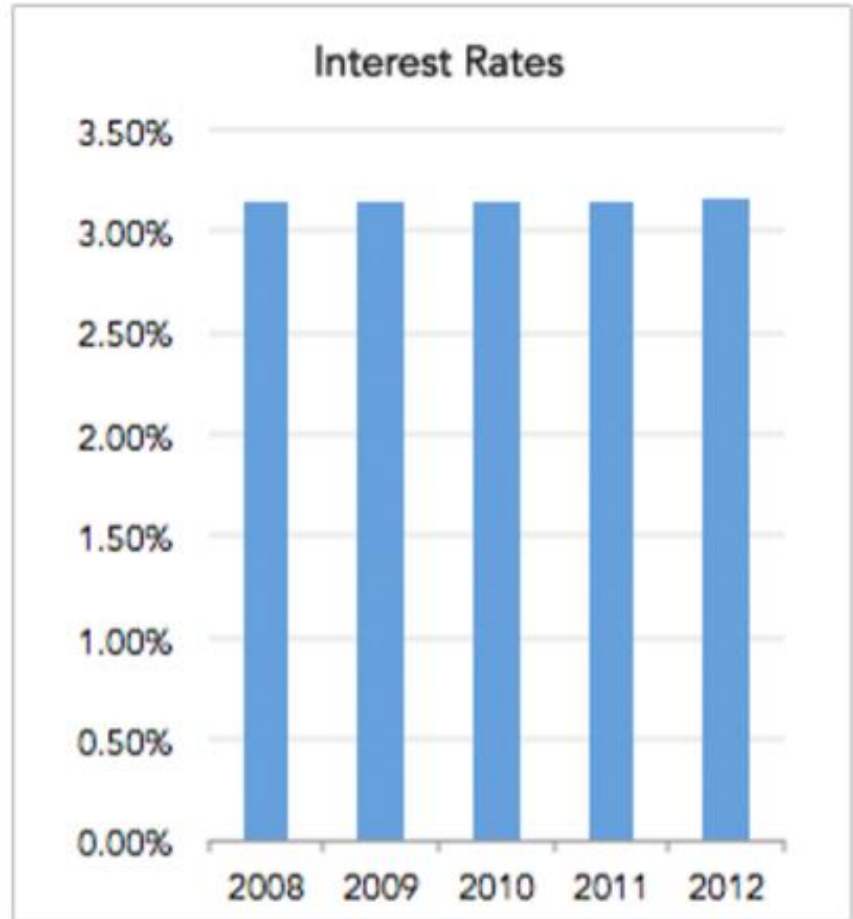
Other than misleading charts, what are other **representation** problems?
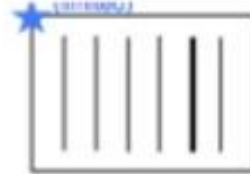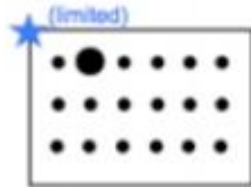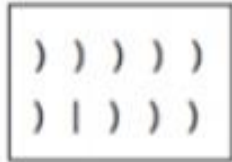
Preattentive Attributes

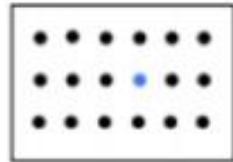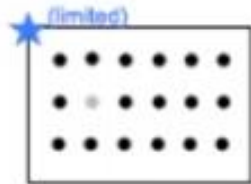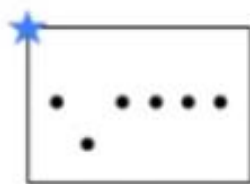Orientation   Shape   Line Length   Line Width
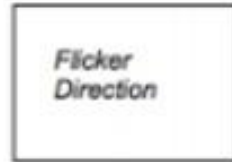
Size   Curvature   Added Marks   Enclosure

Hue   Intensity   2-D Position   Motion

★ Quantitatively perceived
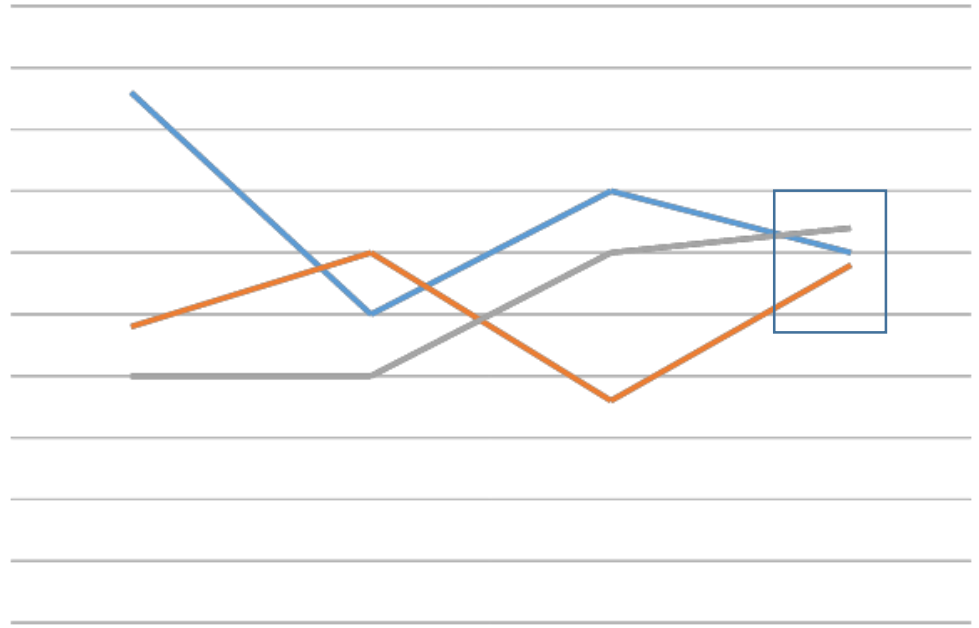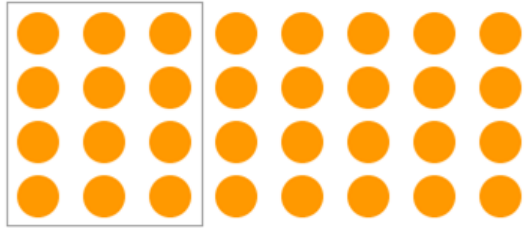
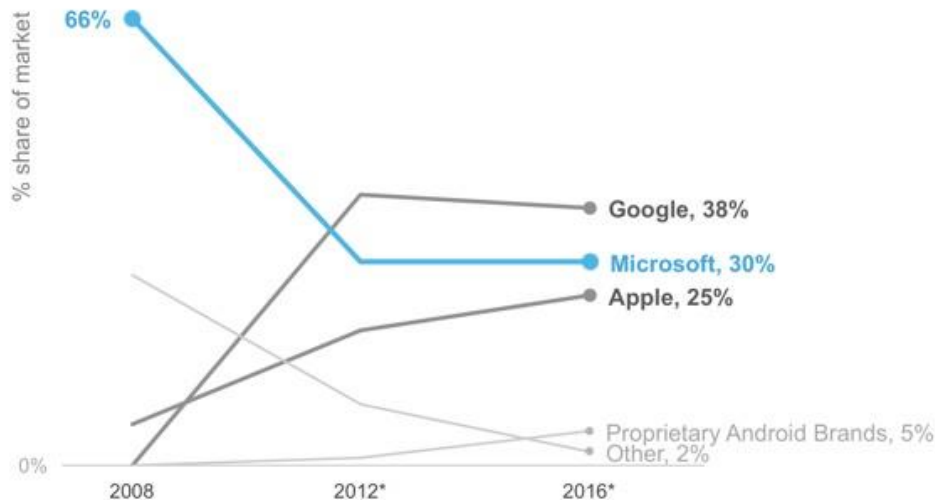# Gestalt Principles of Visual Perception

Choosing Charts

Line

Show trends using a variable of any type with an interval or ratio variable.

Best Practice:

- Use solid lines only

- Don't use too many lines

- Make height so that lines take up roughly 2/3 of chart height.

After spending years dominating the operating system market, **Microsoft is destined to become one of three major players**

Global personal computing device sales by operating system

% share of market

66%

Google, 38%

Microsoft, 30%

Apple, 25%

Proprietary Android Brands, 5%
Other, 2%

0%

2008   2012*   2016*

*Forecasted based on…(I don't know what it's based on, but that detail should be added!)
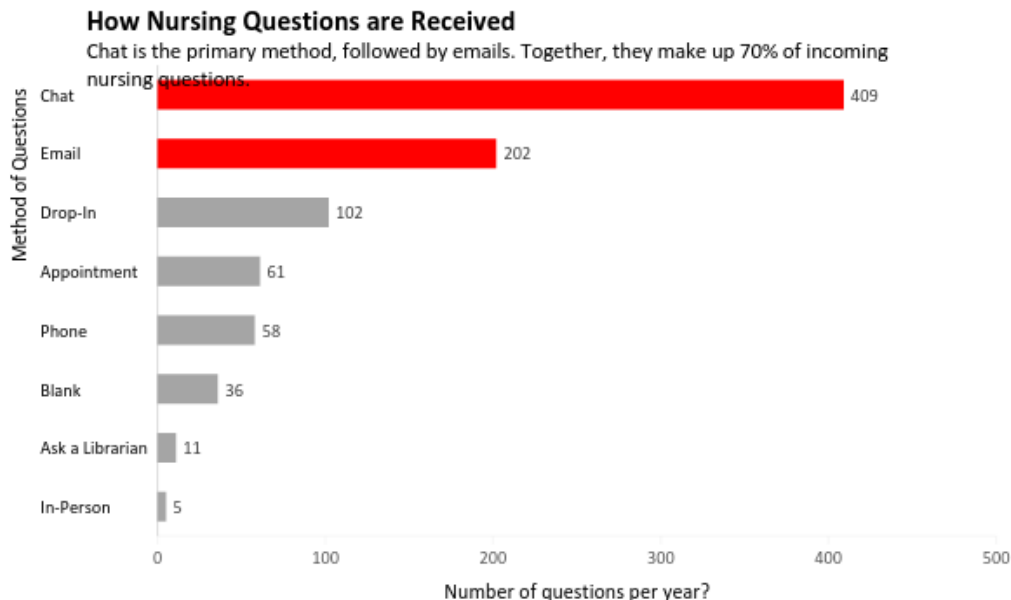
BEST PRACTICES

Horizontal Bar
Show comparison, using a nominal or ordinal variable and an interval or ratio variable.

Best Practice:

- Avoid clutter
- Use to show negative numbers

How Nursing Questions are Received

Chat is the primary method, followed by emails. Together, they make up 70% of incoming nursing questions.

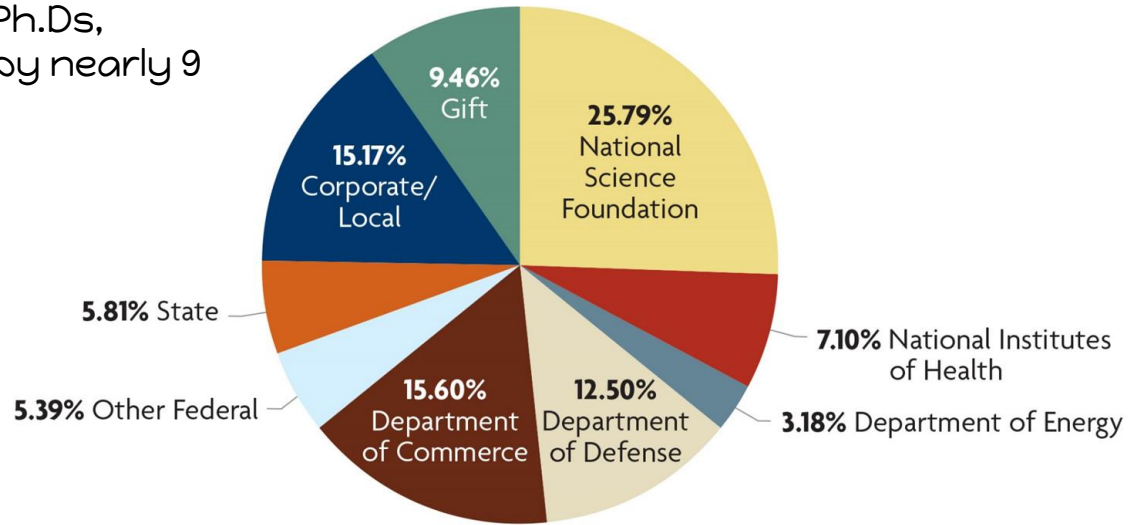| Method of Questions	Number of questions per year?
Chat	409
Email	202
Drop-In	102
Appointment	61
Phone	58
Blank	36
Ask a Librarian	11
In-Person	5

Color used sparingly grabs attention

What's the message?

UTA advances toward Tier 1 status, graduates more than 200 Ph.Ds, increases research expenditures by nearly 9 percent to $85 million.

- Remove clutter
- Push back information that supports but isn't your story

FY 2016 RESEARCH EXPENDITURES

9.46% Gift

15.17% Corporate/Local

25.79% National Science Foundation

5.81% State

7.10% National Institutes of Health

5.39% Other Federal

15.60% Department of Commerce

12.50% Department of Defense

3.18% Department of Energy

http://www.uta.edu/news/releases/2016/09/toward-tier-1.php#sthash.K0XFYdPW.dpuf

Some Changes

FY 2016 Expenditures

- National Science Foundation, 26%
- Department of Energy, 3%
- Other Federal, 5%
- State, 6%
- National Institutes of Health, 7%
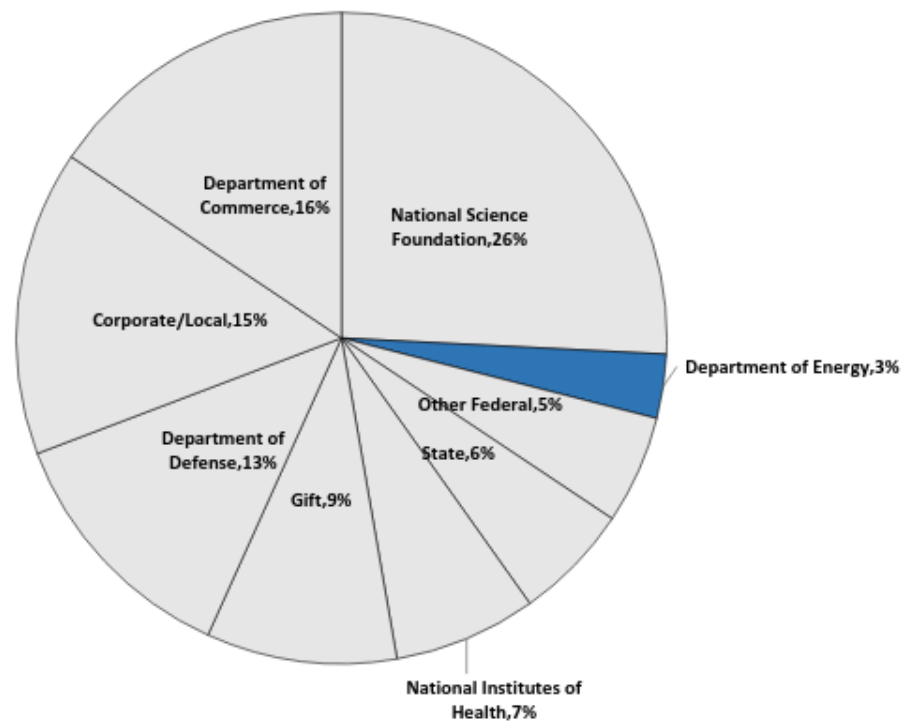- Gift, 9%
- Department of Defense, 13%
- Corporate/Local, 15%
- Department of Commerce, 16%
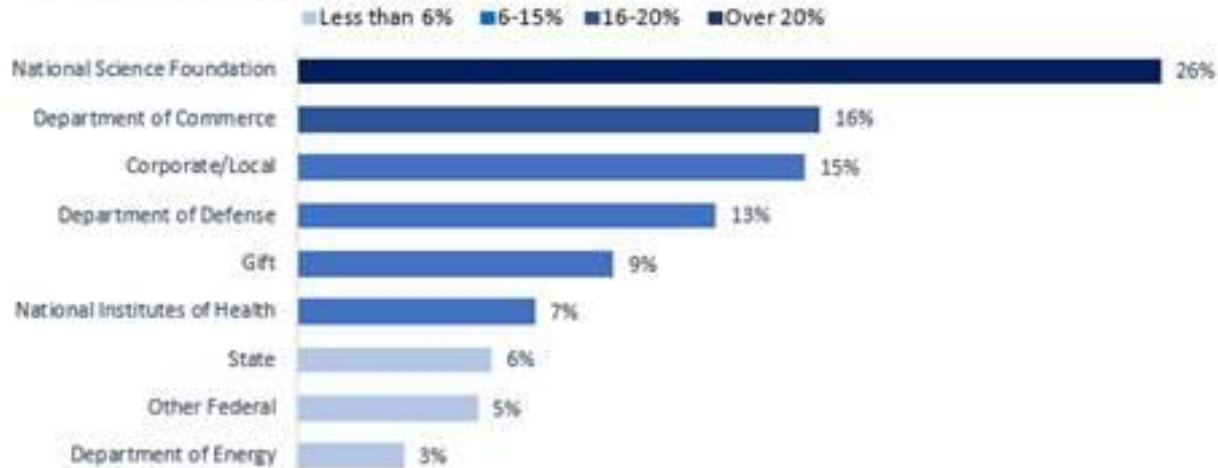
ANOTHER OPTION

FY 2016 - Research Expenditures by Source

UTA generated over $85 million in research expenditures in 2016, up nearly 9% from 2015. Over a fourth of this is through NSF grants.

Legend: Less than 6% | 6-15% | 16-20% | Over 20%

Source	Percentage
National Science Foundation	26%
Department of Commerce	16%
Corporate/Local	15%
Department of Defense	13%
Gift	9%
National Institutes of Health	7%
State	6%
Other Federal	5%
Department of Energy	3%

Other Decisions?

Article Title: UTA advances toward Tier 1 status, graduates more than 200 PhDs, increases research expenditures by nearly 9 percent to $85 million

- Display expenditures
- Display PhD graduates
- Display faculty memberships and awards
- Display UTA-developed patents
- Rate of hiring "star faculty" in comparison with other public universities

http://www.uta.edu/news/releases/2016/09/toward-tier-1.php#sthash.K0XFYdPW.dpuf
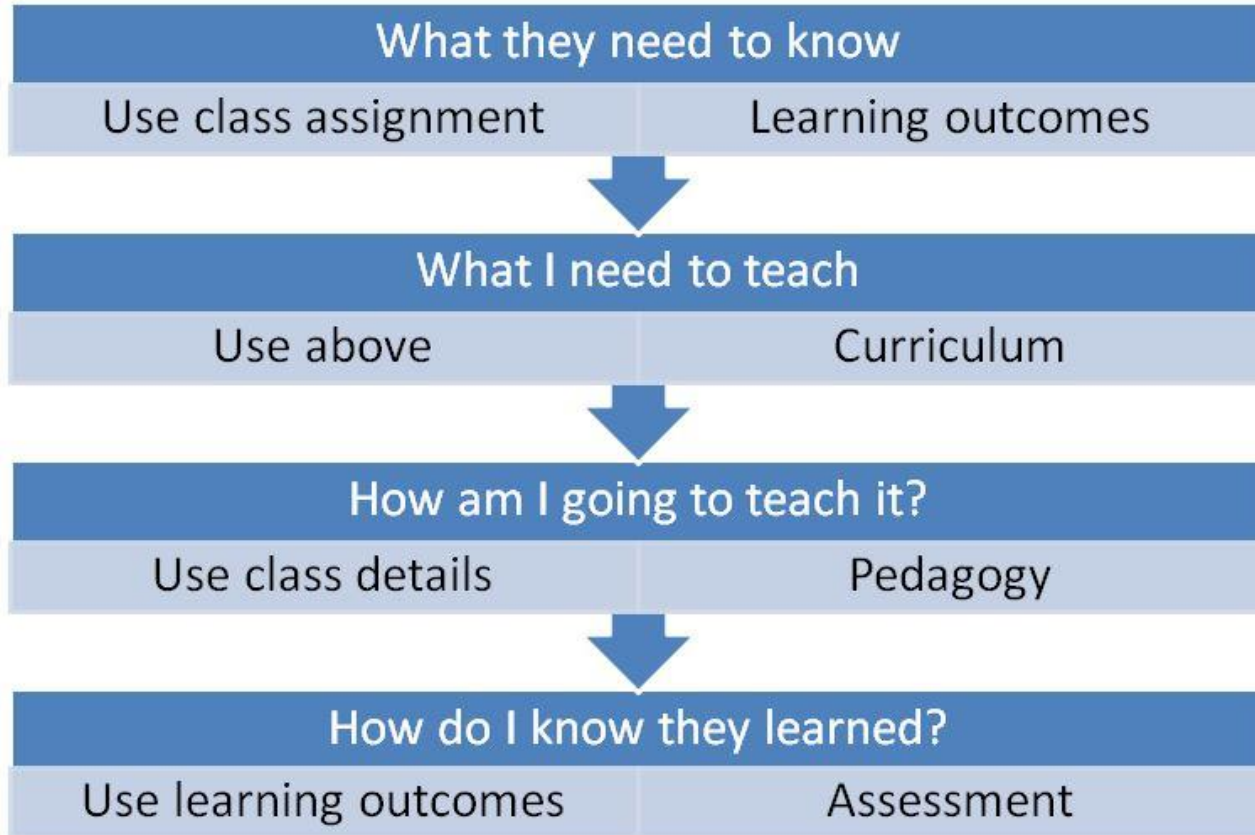
Data Literacy in Classes

The alignment of data literacy competencies with course expectations and assignments.

Context is key!

Data Literacy in Classes

What they need to know
Use class assignment | Learning outcomes

↓

What I need to teach
Use above | Curriculum

↓

How am I going to teach it?
Use class details | Pedagogy

↓

How do I know they learned?
Use learning outcomes | Assessment

Phase 2:
SAMPLE LESSON PLANS: ENGLISH
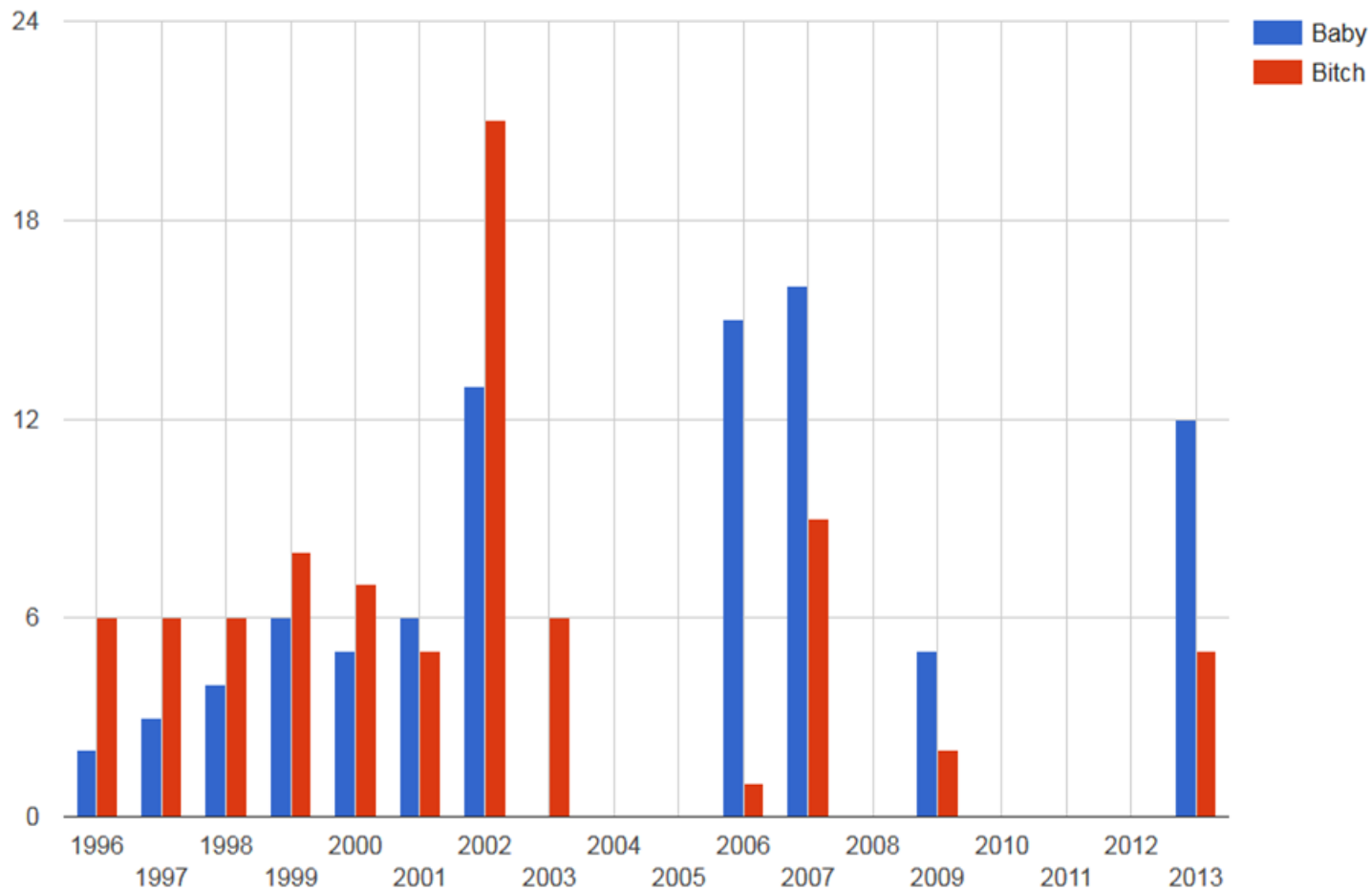
Learning Goals:

- **Understand the use of data in answering research questions**

- Understand the types of metadata and the use of metadata in humanities

- Understand best use of visualizations in answering research questions

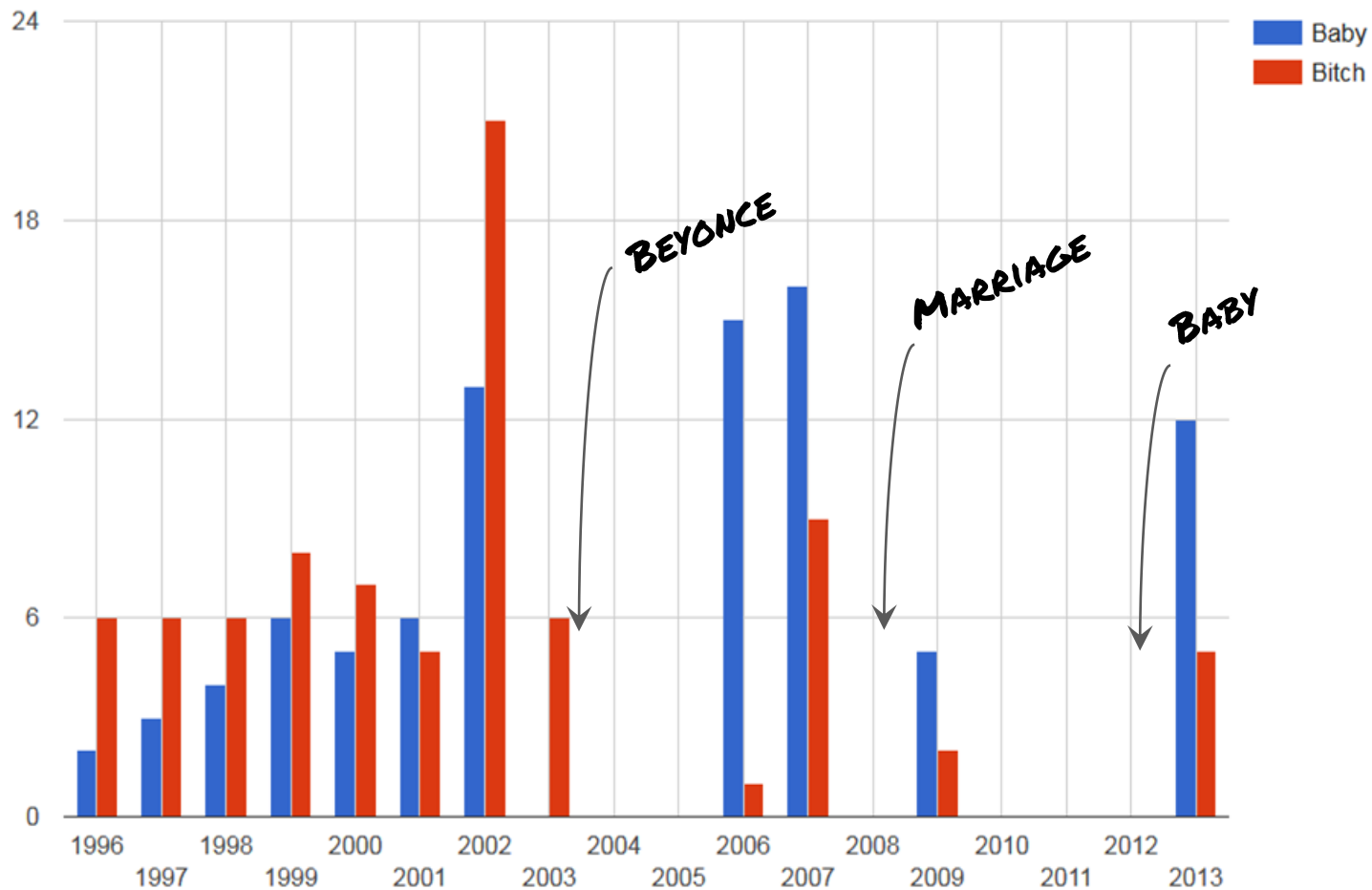- Gain proficiency in use of **Google Fusion Tables** for creating charts

Activities:

- Song lyrics

- Visualization

Tinie Tempah

Jay Z's References Toward Women

Jay Z's References Toward Women

Phase 2:
Sample Lesson Plans: Journalism

Learning Goals:

- Recognize the role of data in research and decision-making

- Recognize areas of potential bias and ambiguous or misleading representation in reporting

- Evaluates the quality of data description and variables available from external sources

Activities:

- Assessing data sources & visualizations

- Discuss data sources and interpretations

Syrian conflict data

CIVILIANS KILLED	40146
OF WHICH:	
WOMEN	NEARLY 4000
CHILDREN	MORE THAN 5800
REBEL FIGHTERS	21,850
REGIME ARMY SOLDIERS	27654
PRO-REGIME MILITIA	17824
HEZBOLLAH	171
UNIDENTIFIED	2726

Syrian Conflict Data

CIVILIANS KILLED		40146
OF WHICH:		
WOMEN	A	NEARLY 4000
CHILDREN		MORE THAN 5800
REBEL FIGHTERS	B	21,850
REGIME ARMY SOLDIERS		27654
PRO-REGIME MILITIA		17824
HEZBOLLAH		171
UNIDENTIFIED		2726

OVERALL LEAST DEROGATORY LANGUAGE, BY STATE

All least derogatory language, per 100,000 tweets

State	Value
Wyoming	120
Montana	121
Vermont	176
South Dakota	192
Idaho	232
Arkansas	244
Minnesota	284
Maine	288
North Dakota	290
Wisconsin	298

Source: Twitter

http://fusion.net/story/277946/hate-filled-cities-and-states-by-tweets/
https://www.abodo.com/blog/tolerance-in-america
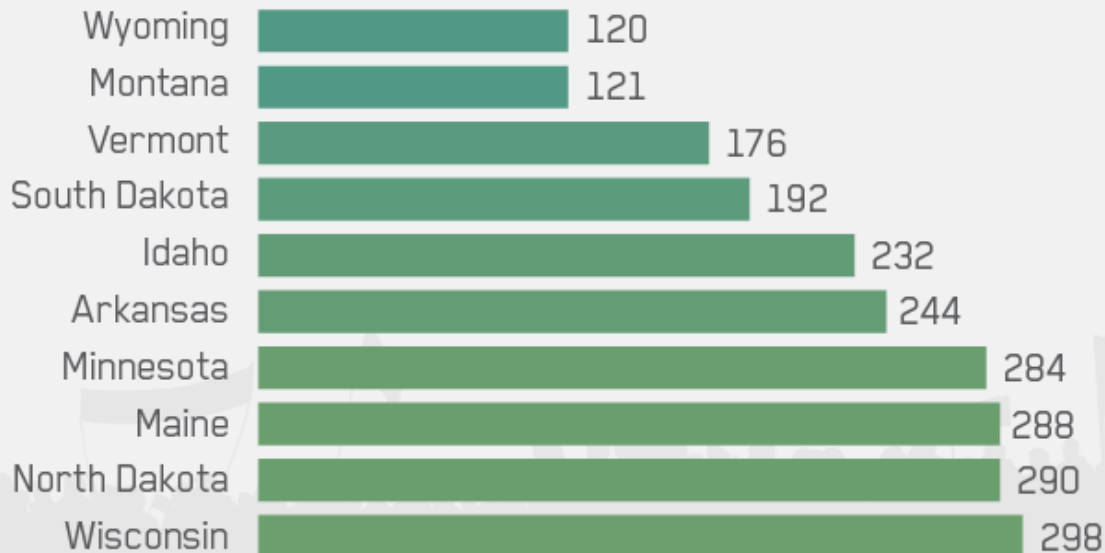
Phase 2:
Sample Lesson Plans: Journalism

- Data management & spreadsheets
  - Data management and organization
  - Data preparation
  - Data structuring and cleaning

- Best practices for visualizations
- Visualization Tools & Tableau

Phase 2:
Sample Lesson Plans: Public Health

Learning Goals:

- Data awareness & knowledge

- Data ethics

- Data responsibility

- Discovery & acquisition of data

- Data management & organization

- Data structuring & cleaning

- Data visualization & representation

Activities:

- IRB training

- Collecting data

- Analyzing and visualizing open data

Website Example

Website Example

ADDICTION AND SUBSTANCE ABUSE
PUBLIC HEALTH ISSUES SUROUNDING OPIOID DRUG ADDICTION

In recent years, drug addiction and substance abuse have come to the forefront of America's consciousness.From popular culture touting the benefits of marijuana to frequent commercials for rehab facilities designed to look like spas, one would be lead to believe that addiction and substance abuse was common place, even normal.

However this is far from the truth. Drug addiction and substance abuse are a burgeoning epidemics within the United States.

(123rf.com, 2016)

From the over prescription of pain medications to easy availability of new synthetic drugs like "bath salts", drug abuse has most certainly become more than just a problem; it's now safe to say that it could be considered endemic in many states, if not the entirety of the U.S.

We don't intend to publicly shame people, but we do want to address the issue of what is often a damaging, risky, and harmful life style.

In this site we will begin to attempt to detail the public health issues that stem from a cycle of users, abusers, and recoverers.

(Denali Healthcare, 2014)

Website Example

Prescription Drugs and Addiction

When you think of drug addiction, what's the first thing that comes to mind? The majority of us would immediately say that substances like heroin, cocaine, and even to some extent marijuana are the main causes of the problem.

While these drugs are well known to that contribute to the problem of addiction in the United States there are unfortunately other substances that are just as problematic. In recent years we've been constantly bombarded with tales of doctors across the nation over prescribing opioids. It's this easy access that can lead to misuse, addiction, and in some cases even the death of the patients, these deaths have reached epidemic levels within the past decade.

From 2000 to 2010, the rates of unintentional overdose on prescription opioids increased dramatically, accounting for more deaths than those attributed to all illicit drug categories combined (Washington, 2016)(CDC Vital Signs, 2011). Drug addiction is an issue that is growing with severity every day and can be attributed to over prescription of opioids, loose regulation guidelines from the U.S Food and Drug Administration agency (FDA) and miscommunication between health professionals.

In West Virginia, a place considered to be at the epicenter of the opioid epidemic, Dr. Michael Kostenko currently stands under investigation, and is threatened with loss of his license after multiple reported deaths caused by his practice of over prescribing opioids to his patients. Dr. Kostenko is estimated to have written well over 40,000 prescriptions for opioid substances like Oxycodone, Vicodin, or Oxymorphone within the past two years. When interviewed, he was asked, "Did you write 325 prescriptions the first week of January for more than 19,000 oxycodone pills?" Dr. Kostenko replied "Possibly, it may well be." (Axelrod, 2016) Unfortunately, Dr.Kostenko's case is not a unique one. In West Virginia alone, seven doctors have lost their licenses and 15 more are currently being investigated. (Axelrod, 2016) In the United States as a whole, it was estimated in 2014 that U.S doctors wrote nearly 200 million prescriptions for opioid painkillers, while roughly 19,000 deaths linked to the drugs were reported.(Ballantyne, 2015)

While these numbers are alarming to hear, the bigger question is why are we as Americans so over prescribed painkillers? A clue is that some medical professionals,

PRESCRIPTION PAINKILLERS

Prescription painkillers often refers to opioid or narcotic pain relievers, including drugs such as Vicodin (hydrocodone), OxyContin (oxycodone), Opana (oxymorphone), and methadone. (CDC Vital Signs, 2011)

15,000

Nearly 15,000 people die every year of overdoses involving prescription painkillers. (CDC Vital Signs, 2011)
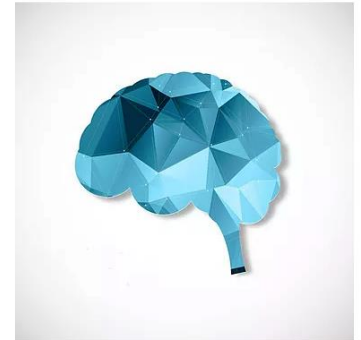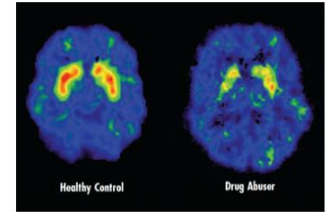
Website Example

- A chronic, relapsing, brain disease known by it's compulsive drug seeking and use, despite the harmful consequences to the user.
- Considered a brain disease because drugs change the brain – they change it's structure and how it works.

ADDICTION & YOUR BRAIN
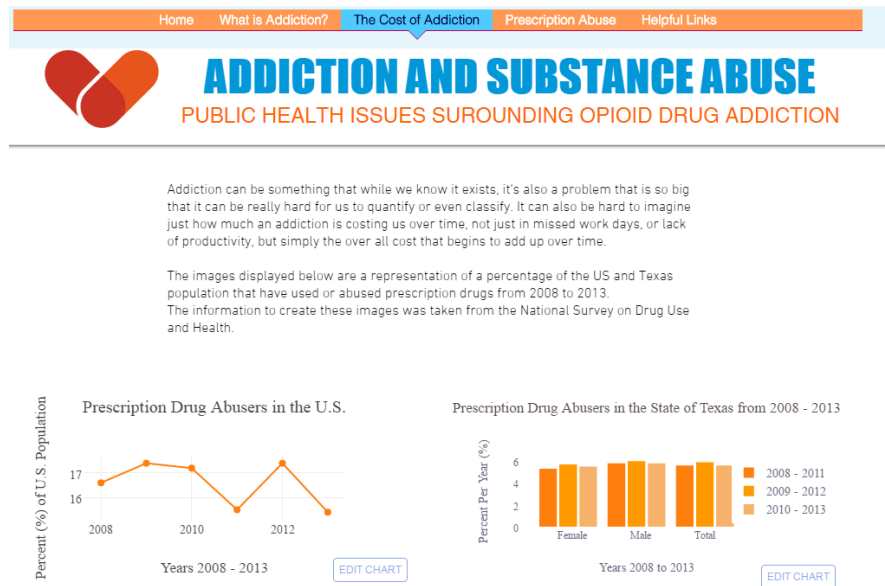
Drugs effect 3 primary areas of our brain –

- The brain stem
  - In charge of all the functions we need it to stay alive —breathing, moving blood, and digesting food. It also links the brain with the spinal cord, which runs down the back and moves muscles and limbs as well. It lets the brain know what's happening to the body.
- The limbic system
  - It links together a bunch of brain structures that control our emotional responses, such as feeling pleasure when we eat chocolate. These good feelings motivate us to repeat the behavior, good or bad.
- The cerebral cortex
  - A mushroom-shaped outer part of the brain (the gray matter). In us humans, it is so big that it makes up about three-fourths of our entire brain. It's divided into four areas, which are called lobes, that control specific functions. Some areas process information from our senses; allowing us to see, feel, hear, and taste. The front part of the cortex, is known as the frontal cortex or our forebrain, it's our thinking center. It powers our ability to think, plan, solve problems, and make decisions (good or bad).

- National Institute for Drug Abuse for Teens

(Harryarts - Freepik.com)

Website Example

DATA ANALYSIS & VISUALIZATION

UTA Libraries workshop series focused on data visualization, GIS, and statistical tools.

UTA Libraries' Division of Scholarly Communication
library.uta.edu/scholcomm

Open Workshops
libguides.uta.edu/DAVis

Data management

Data visualization

Tools (ArcGIS, SPSS)

ABCs of DMPs

Open data & data publishing

What's Next?

Data Literacy Competencies

- More campus integration

- Data portal

- Research on the competencies' effectiveness at other campuses

- Data CAVE

- Data repository

- Data training

Up Next...
Laying the Foundation for Research Data Services

Session 3 – Learning By Example: Connecting Data Competencies with the Texas Data Repository

- ◦ Thursday, March 9, 2017 | 11:00 AM - 12:00 PM (CST)
- ◦ This webinar session will connect teaching data competencies explored in Session 2 with core functionality of the Texas Data Repository. It will address data competency learning outcomes, including: using the TDR to add, describe, share, and publish data; managing data, including versioning and de-accessioning data; and downloading and using data.

Texas Digital Library

Questions and Discussion