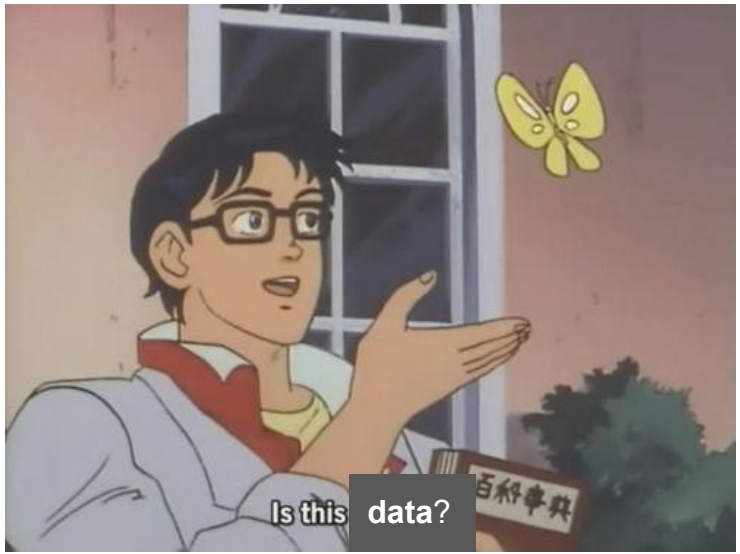
Data Literacy 101:

For Water Conservation Professionals

CALIFORNIA
DATA
COLLABORATIVE

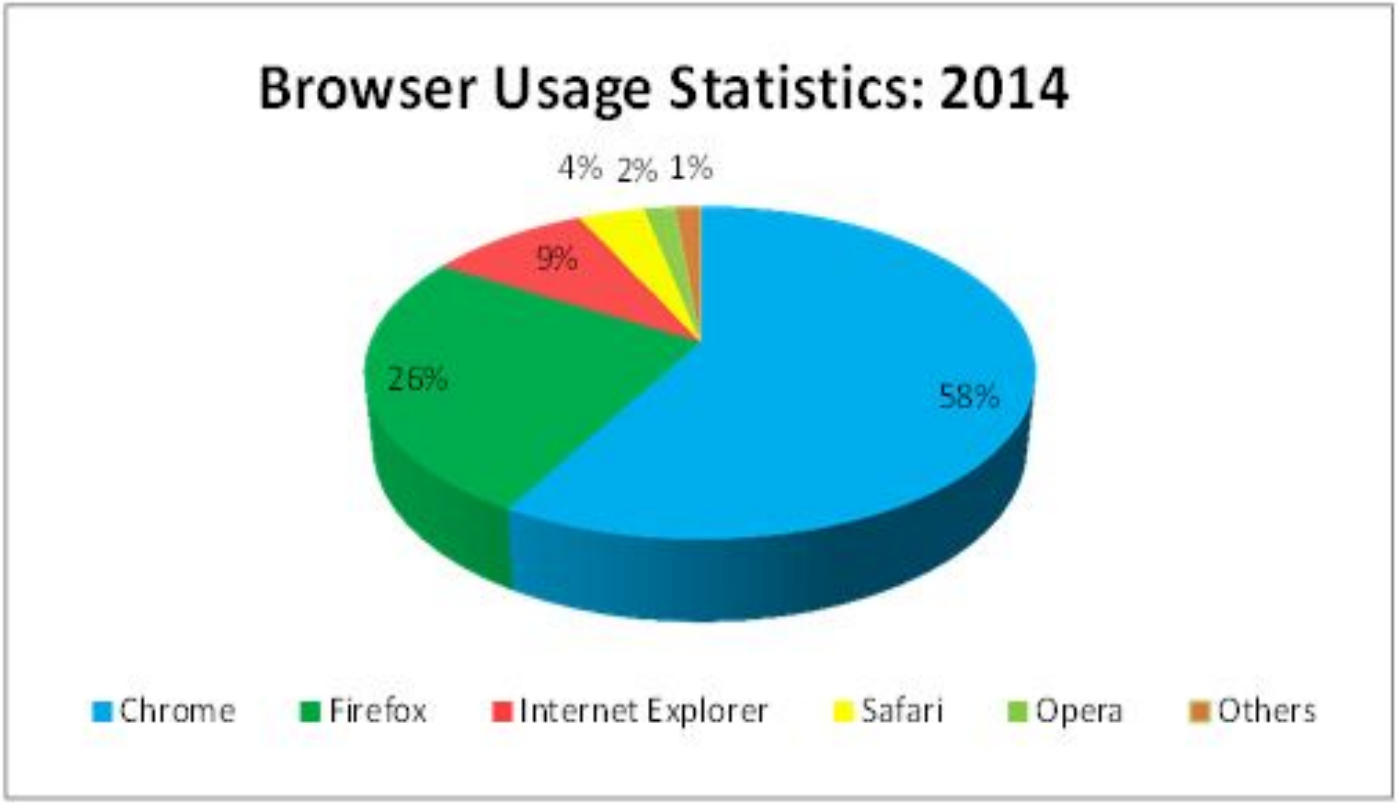
CALIFORNIA
WATER EFFICIENCY
PARTNERSHIP

Warm-up: Is it data?



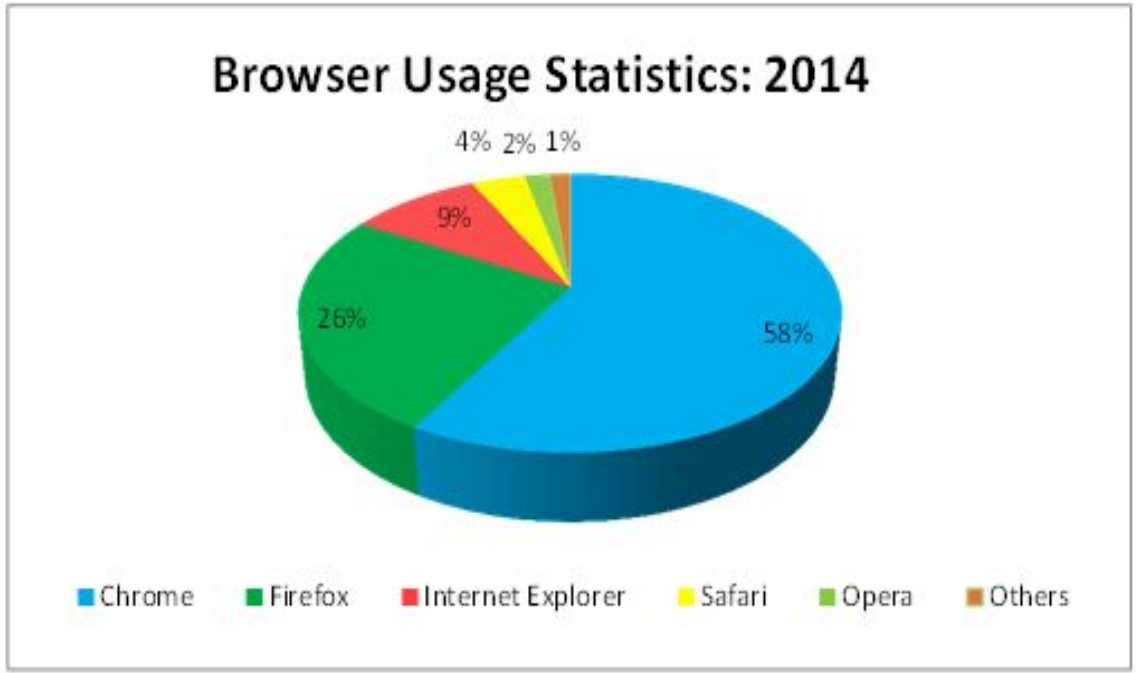
- 4 examples
- In the chat - reply “Yes” or “No” and any justification
- Think about the reason **why** you selected what you did

Is it data?


Browser Usage Statistics: 2014
4%
2%
1%
9%
26%
58%
Chrome
Firefox
Internet Explorer
Safari
Opera
Others

Is it data?

NO: This is a data visualization (*pie chart*)



- What is the total that each percentage is derived from?
- How many “Others” are there, and what are they?

Is it data?

Monthly Cash Flow			
	Actual	Budget	Variance
Cash received			
Fees	$21,571	$20,000	$1,571
Salary grants	10,005	11,000	(995)
Other	76		76
	31,652	31,000	652
Cash paid out			
Salaries and benefits	21,575	20,000	(1,575)
Food	2,350	2,000	(350)
Play supplies	335	500	165
Other	3,270	1,500	(1,770)
	27,530	24,000	(3,530)
Excess of cash received over cash paid out	$4,122	$7,000	$(2,878)

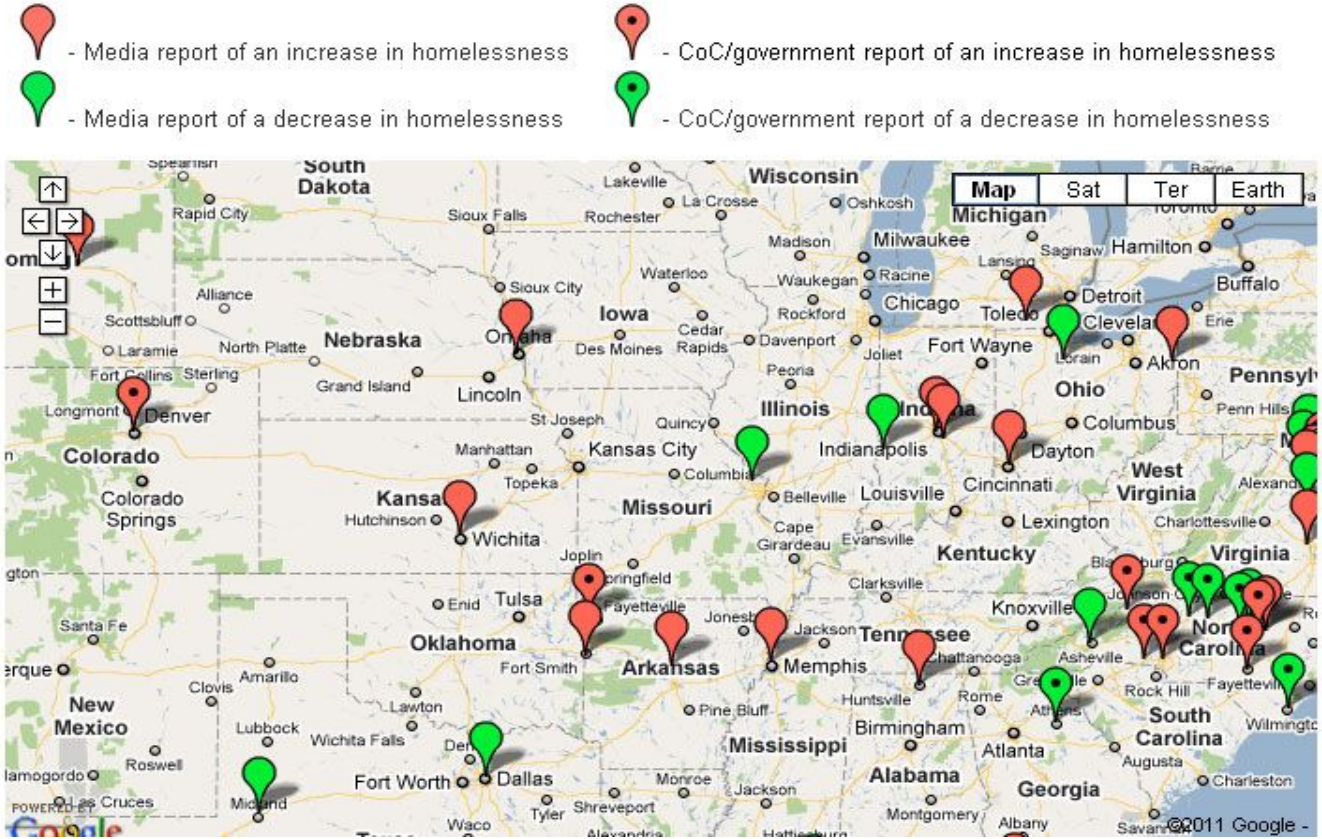
Is it data?

NO: This is a report

Monthly Cash Flow	Actual	Budget	Variance
Cash received			
Fees	$21,571	$20,000	$1,571
Salary grants	10,005	11,000	(995)
Other	76		76
	31,652	31,000	652
Cash paid out			
Salaries and benefits	21,575	20,000	(1,575)
Food	2,350	2,000	(350)
Play supplies	335	500	165
Other	3,270	1,500	(1,770)
	27,530	24,000	(3,530)
Excess of cash received over cash paid out	$4,122	$7,000	$(2,878)

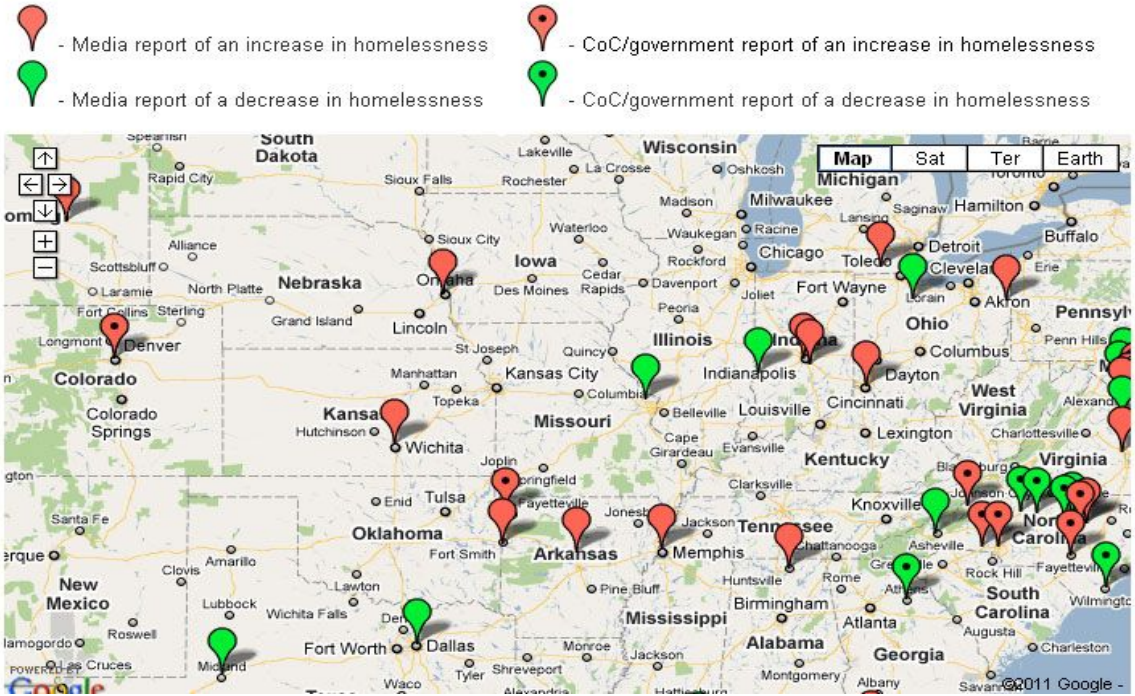
- How does this month compare to last month?
- Which category is the most over-budget?

Is it data?


- Media report of an increase in homelessness
- CoC/government report of an increase in homelessness
- Media report of a decrease in homelessness
- CoC/government report of a decrease in homelessness
Map
Sat
Ter
Earth
South Dakota
Wisconsin
Michigan
Nebraska
Iowa
Illinois
Ohio
Colorado
Kansas
Missouri
Oklahoma
Arkansas
Kentucky
Tennessee
Mississippi
Alabama
Georgia
New Mexico
West Virginia
Pennsylvania
Virginia
South Carolina
Rapid City
Sioux Falls
Rochester
La Crosse
Oshkosh
Milwaukee
Madison
Chicago
Detroit
Toledo
Cleveland
Buffalo
Akron
Fort Wayne
Indianapolis
Columbus
Dayton
Cincinnati
Louisville
Lexington
Charleston
Sioux City
Omaha
Des Moines
Cedar Rapids
Davenport
Rockford
Waterloo
Peoria
Joliet
Lincoln
Grand Island
North Platte
Scottsbluff
Alliance
Laramie
Fort Collins
Sterling
Longmont
Denver
Colorado Springs
St Joseph
Quincy
Kansas City
Manhattan
Topeka
Columbia
Belleville
Hutchinson
Wichita
Joplin
Springfield
Cape Girardeau
Evansville
Clarksville
Enid
Tulsa
Fayetteville
Fort Smith
Jonesboro
Jackson
Memphis
Knoxville
Chattanooga
Asheville
Rock Hill
Santa Fe
Amarillo
Clovis
Lawton
Wichita Falls
Lubbock
Roswell
Pine Bluff
Huntsville
Rome
Birmingham
Atlanta
Augusta
Charleston
Fort Worth
Dallas
Midland
Monroe
Jackson
Shreveport
Tyler
Waco
Montgomery
Albany
©2011 Google -

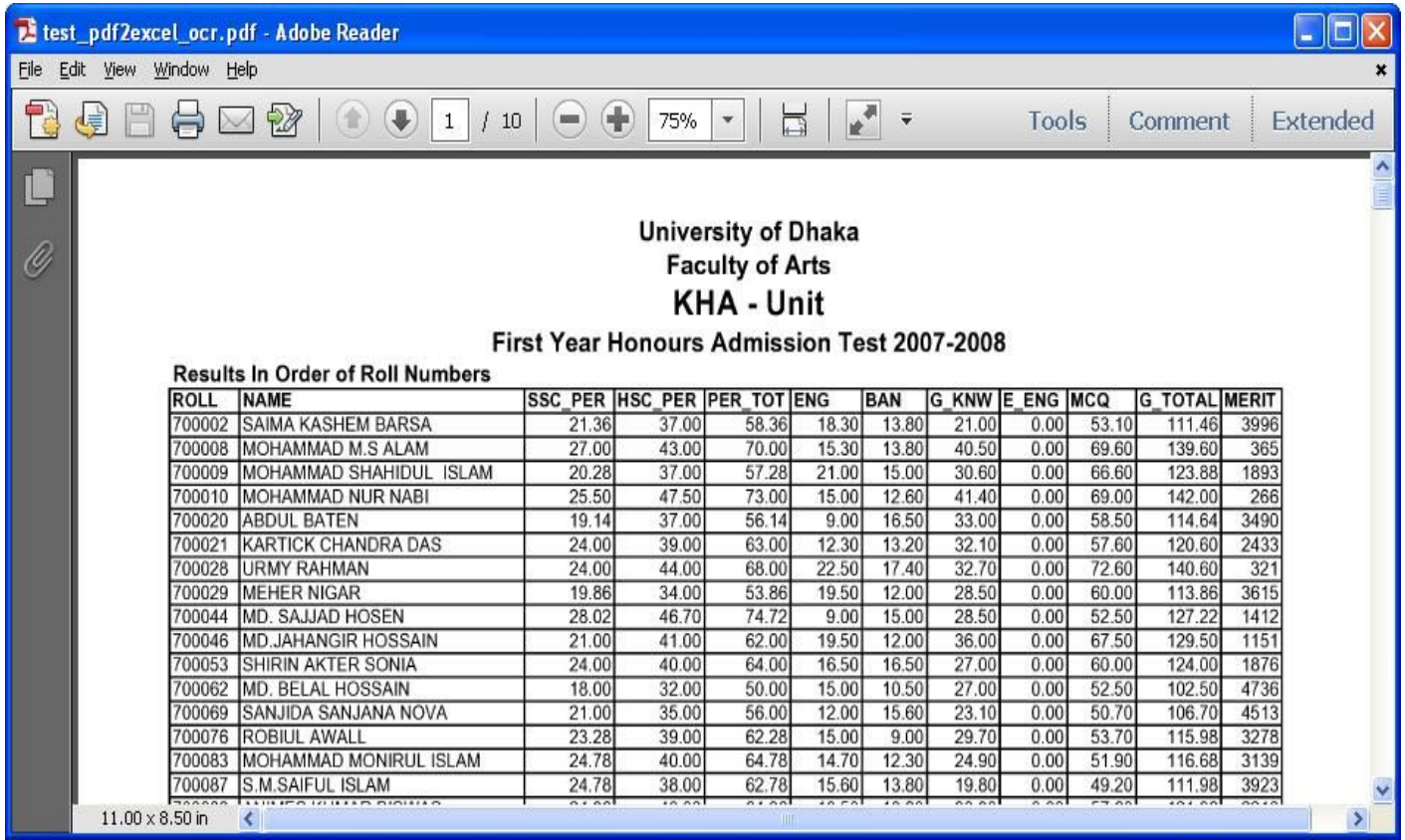
Is it data?

NO: This is a visualization/map



- Which regions/states are seeing the most change?
- What does a “report” mean? Are there numbers to back it up?

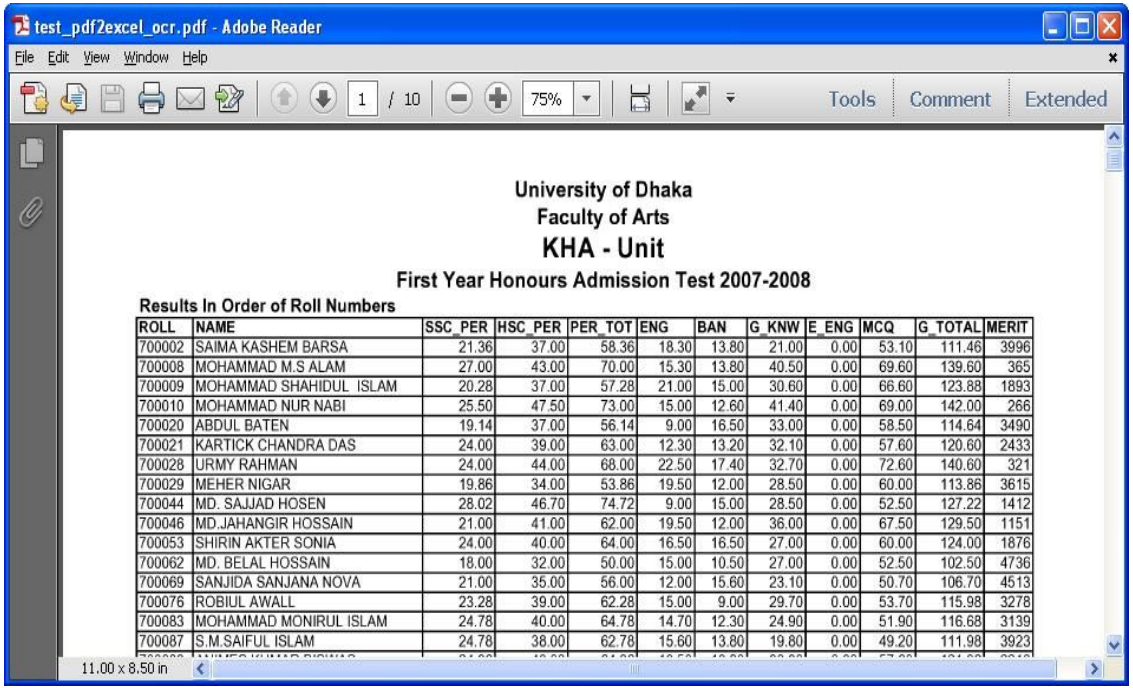
Is it data?



ROLL	NAME	SSC_PER	HSC_PER	PER_TOT	ENG	BAN	G_KNW	E_ENG	MCQ	G_TOTAL	MERIT
700002	SAIMA KASHEM BARSA	21.36	37.00	58.36	18.30	13.80	21.00	0.00	53.10	111.46	3996
700008	MOHAMMAD M.S ALAM	27.00	43.00	70.00	15.30	13.80	40.50	0.00	69.60	139.60	365
700009	MOHAMMAD SHAHIDUL ISLAM	20.28	37.00	57.28	21.00	15.00	30.60	0.00	66.60	123.88	1893
700010	MOHAMMAD NUR NABI	25.50	47.50	73.00	15.00	12.60	41.40	0.00	69.00	142.00	266
700020	ABDUL BATEN	19.14	37.00	56.14	9.00	16.50	33.00	0.00	58.50	114.64	3490
700021	KARTICK CHANDRA DAS	24.00	39.00	63.00	12.30	13.20	32.10	0.00	57.60	120.60	2433
700028	URMY RAHMAN	24.00	44.00	68.00	22.50	17.40	32.70	0.00	72.60	140.60	321
700029	MEHER NIGAR	19.86	34.00	53.86	19.50	12.00	28.50	0.00	60.00	113.86	3615
700044	MD. SAJJAD HOSEN	28.02	46.70	74.72	9.00	15.00	28.50	0.00	52.50	127.22	1412
700046	MD.JAHANGIR HOSSAIN	21.00	41.00	62.00	19.50	12.00	36.00	0.00	67.50	129.50	1151
700053	SHIRIN AKTER SONIA	24.00	40.00	64.00	16.50	16.50	27.00	0.00	60.00	124.00	1876
700062	MD. BELAL HOSSAIN	18.00	32.00	50.00	15.00	10.50	27.00	0.00	52.50	102.50	4736
700069	SANJIDA SANJANA NOVA	21.00	35.00	56.00	12.00	15.60	23.10	0.00	50.70	106.70	4513
700076	ROBIUL AWALL	23.28	39.00	62.28	15.00	9.00	29.70	0.00	53.70	115.98	3278
700083	MOHAMMAD MONIRUL ISLAM	24.78	40.00	64.78	14.70	12.30	24.90	0.00	51.90	116.68	3139
700087	S.M.SAIFUL ISLAM	24.78	38.00	62.78	15.60	13.80	19.80	0.00	49.20	111.98	3923

Is it data?

NO: This is a PDF



University of Dhaka
Faculty of Arts
KHA - Unit
First Year Honours Admission Test 2007-2008

Results In Order of Roll Numbers

ROLL	NAME	SSC_PER	HSC_PER	PER_TOT	ENG	BAN	G_KNW	E_ENG	MCQ	G_TOTAL	MERIT
700002	SAIMA KASHEM BARSA	21.36	37.00	58.36	18.30	13.80	21.00	0.00	53.10	111.46	3996
700008	MOHAMMAD M.S ALAM	27.00	43.00	70.00	15.30	13.80	40.50	0.00	69.60	139.60	365
700009	MOHAMMAD SHAHIDUL ISLAM	20.28	37.00	57.28	21.00	15.00	30.60	0.00	66.60	123.88	1893
700010	MOHAMMAD NUR NABI	25.50	47.50	73.00	15.00	12.60	41.40	0.00	69.00	142.00	266
700020	ABDUL BATEN	19.14	37.00	56.14	9.00	16.50	33.00	0.00	58.50	114.64	3490
700021	KARTICK CHANDRA DAS	24.00	39.00	63.00	12.30	13.20	32.10	0.00	57.60	120.60	2433
700028	URMY RAHMAN	24.00	44.00	68.00	22.50	17.40	32.70	0.00	72.60	140.60	321
700029	MEHER NIGAR	19.86	34.00	53.86	19.50	12.00	28.50	0.00	60.00	113.86	3615
700044	MD. SAJJAD HOSEN	28.02	46.70	74.72	9.00	15.00	28.50	0.00	52.50	127.22	1412
700046	MD.JAHANGIR HOSSAIN	21.00	41.00	62.00	19.50	12.00	36.00	0.00	67.50	129.50	1151
700053	SHIRIN AKTER SONIA	24.00	40.00	64.00	16.50	16.50	27.00	0.00	60.00	124.00	1876
700062	MD. BELAL HOSSAIN	18.00	32.00	50.00	15.00	10.50	27.00	0.00	52.50	102.50	4736
700069	SANJIDA SANJANA NOVA	21.00	35.00	56.00	12.00	15.60	23.10	0.00	50.70	106.70	4513
700076	ROBIUL AWALL	23.28	39.00	62.28	15.00	9.00	29.70	0.00	53.70	115.98	3278
700083	MOHAMMAD MONIRUL ISLAM	24.78	40.00	64.78	14.70	12.30	24.90	0.00	51.90	116.68	3139
700087	S.M.SAIFUL ISLAM	24.78	38.00	62.78	15.60	13.80	19.80	0.00	49.20	111.98	3923

- What is the average MERIT score?
- Who had the highest overall performance?

Takeaways

- All of these examples are created from **raw material** - a bunch of letters and numbers in a file somewhere. That's the data.
- The real data should be easy to slice and dice to ask new questions
- Think about the data behind the chart, map, or report. What does the rest of it look like? Why are you only being shown this subset?

Some data about me

- Christopher Tull
- Born in Ventura, CA. Live in Oxnard CA.
- Studied Computer Science and Mathematics at CSU Channel Islands
- Spent a year in NYC learning how to use data to help cities operate more efficiently
- Now project manager for the California Data Collaborative

About the California Data Collaborative

- Formed in 2016 as a **nonprofit** coalition of water agencies and has grown into a trusted presence in California's water sector

- Staffed by a team of data and design experts, and governed by a steering committee of water managers

Today we will talk about:

- The Universe of Water Agency Data
- Data Types and Formats
- Accessing and Sharing Data
- Metadata
- Data Terms Kahoot!
- Potential and Pitfalls
- An Applied Example

The Universe of Water Agency Data

Agency

Budgets Account Balances Employee Details Emails IT System Logs Contracts

Interaction between Agency and Customer

Bills Call Notes Rebates
Field Visits Bill Adjustments
Fees Shutoffs Bill Inserts
Marketing Programs Classes

Interaction between Agency and System

Meter Reads New Connections
Capital Improvements Repairs
Replacements Inspections
Maintenance

Customers

Name Address Household Size
Landscape Billing Preferences
Customer Class Dates of Service

System

Area Served Meter Characteristics
Pipe Details Climate Evapotranspiration
Facility Details Operating Specs

The Universe of Water Agency Data

Agency

Budgets Account Balances Employee Details Emails IT System Logs Contracts

Interaction between Agency and Customer

Bills Call Notes Rebates
Field Visits Bill Adjustments
Fees Shutoffs Bill Inserts
Marketing Programs Classes

Interaction between Agency and System

Meter Reads New Connections
Capital Improvements Repairs
Replacements Inspections
Maintenance

Customers

Name Address Household Size
Landscape Billing Preferences
Customer Class Dates of Service

System

Area Served Meter Characteristics
Pipe Details Climate Evapotranspiration
Facility Details Operating Specs

Water Efficiency Data

Name

Billed Use

Call Notes

Customer Class

Year Built

Rebates

Landscape Area

Evapotranspiration

Quantitative Data

Name

Billed Use **(7 CCF)**

Call Notes

Customer Class

Year Built **(1960)**

Rebate **Amount ($250)**

Landscape Area **(2000 Sq.Ft)**

Evapotranspiration **(57 in.)**

Qualitative Data

Name **(Kamala Harris)**

Billed Use

Call Notes **(“Asked about a water audit…”)**

Customer Class **(Single Family)**

Year Built

Rebate **Type (WBIC)**

Landscape Area

Evapotranspiration

Categorical (Qualitative) Data

Name

Billed Use

Call Notes

Customer Class
(Single Family)

Year Built

Rebate **Type**
(WBIC)

Landscape Area

Evapotranspiration

Ordinal (Qualitative) Data

Name

Billed Use

Call Notes

Customer Class

Year Built

Rebate **Date**
(4/22/2021) *

Landscape Area

Evapotranspiration

***(sort of…)**

Data formats

Easy for people to understand vs. easy for computers to understand (“machine readable”)

Common machine readable data types:

FORMAT	OFTEN USED FOR	EXAMPLE
Shapefile (.shp), Geojson (.json)	Spatial data (points, lines, polygons)	Parcel boundaries
Comma Separated Values (.csv), Excel (.xlsx)	Tabular data, spatial data in point form (lat/lng)	Records of rebate participation
Text (.txt)	Textual data	Call notes of customer interactions
Raster (.jpg, .png, .tif)	Images	Aerial photograph
Semi-structured (.json, .yaml)	Sending information over the internet	CIMIS API

customers.csv

account, class, use_ccf, landscape_area

1,Single Family,6,1500

2,Single Family,17,2300

3,Multi Family,274,5000

4,Commercial,153,1300

5,Irrigation,55,9000

6,Irrigation,825,23000

customers.json


```
[
 {
   "account": 1,
   "class": "Single Family",
   "use_ccf": 6,
   "landscape_area": 1500
 },
 {
   "account": 2,
   "class": "Single Family",
…
 }
]
```

customers.shp

Spatial files encode the coordinates of points, lines, and polygons.

Shapefiles are a binary format - need the right software to make sense of it or it ends up like this:

NUL NUL'
NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL
NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL NUL
SOH NUL NUL NUL ýµ SUB ð0WAáê—Qï9>A NUL NUL NUL STX NUL NUL NUL
SOH NUL NUL NUL;
ÆTY3WA¬FUCDD>A NUL NUL NUL ETX NUL NUL NUL
SOH NUL NUL NUL à
US Ú2WAýü;c-E>A NUL NUL NUL EOT NUL NUL NUL
SOH NUL NUL NUL •ll ª ã1WAmY"aSD>A NUL NUL NUL ENQ NUL NUL NUL
SOH NUL NUL NUL (-;ÿ×0WAð•,Ûþ@>A NUL NUL NUL ACK NUL NUL NUL
SOH NUL NUL NUL Ð1U$Ú0WAD)øß÷@>A NUL NUL NUL BEL NUL NUL NUL
SOH NUL NUL NUL -EÍèï0WAÍ~¡á SOH A>A NUL NUL NUL BS NUL NUL NUL
SOH NUL NUL NUL ÓÁ2±%1WA FF W CAN 9¬A>A NUL NUL NUL NUL NUL NUL
SOH NUL NUL NUL kÇ/'1WAoÉk8±A>A NUL NUL NUL
NUL NUL NUL
SOH NUL NUL NUL ð-Zïk1WA Yã>,B>A NUL NUL NUL V† NUL NUL NUL
SOH NUL NUL NUL °¼G1WA/ ESC fÏ SI C>A NUL NUL NUL FF NUL NUL NUL
SOH NUL NUL NUL -3u 1WAŸïrÁ DLE C>A NUL NUL NUL
NUL NUL NUL
SOH NUL NUL NUL ,c- RS Ê3WAÀ ª Œ_+?>A NUL NUL NUL SO NUL NUL NUL

customers.geojson

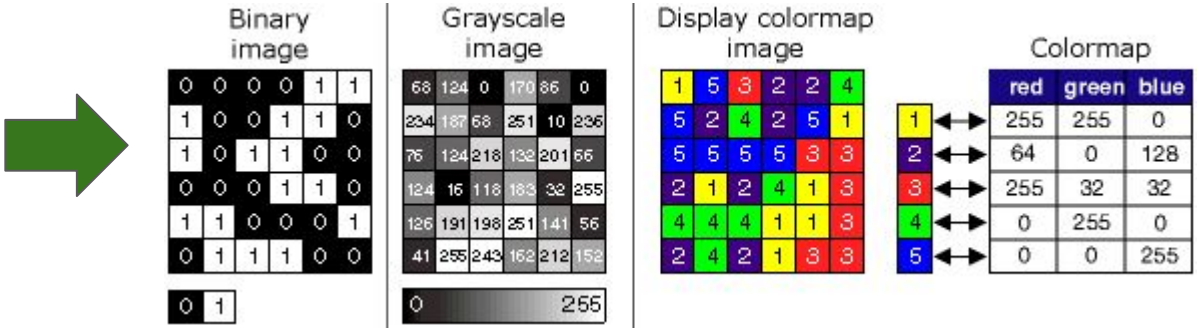
Spatial files encode the coordinates of points, lines, and polygons.

Geojson is an open format based on JSON

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [125.6, 10.1]
  },
  "properties": {
    "name": "Dinagat Islands"
  }
}
```

customers.tif

Raster files encode data in a grid of values

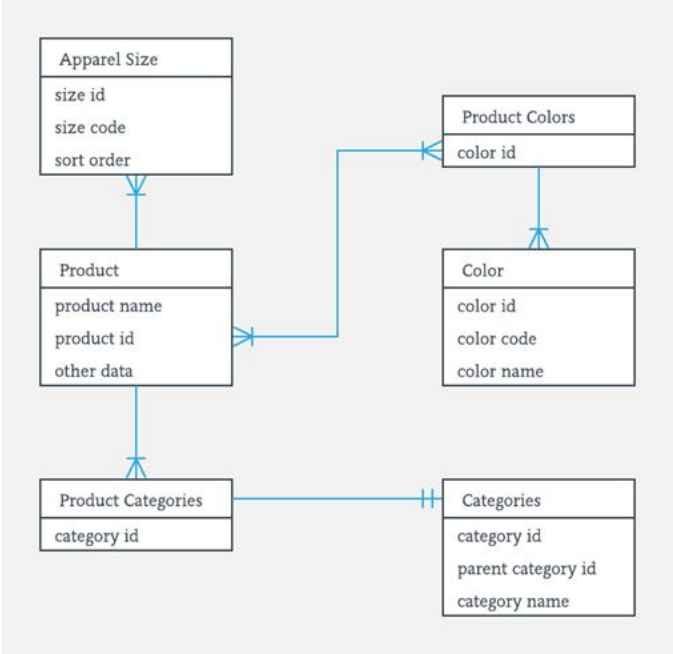


A note on databases

"an organized collection of data, generally stored and accessed electronically from a computer system."

A **relational database** is often used for more permanent data storage. Has nice properties to ensure data consistency and allows querying data with Structured Query Language (SQL)

*SELECT **
FROM product
WHERE product_name = 'TShirt'



Data Openness

Open data vs. closed data

Data availability is a spectrum!

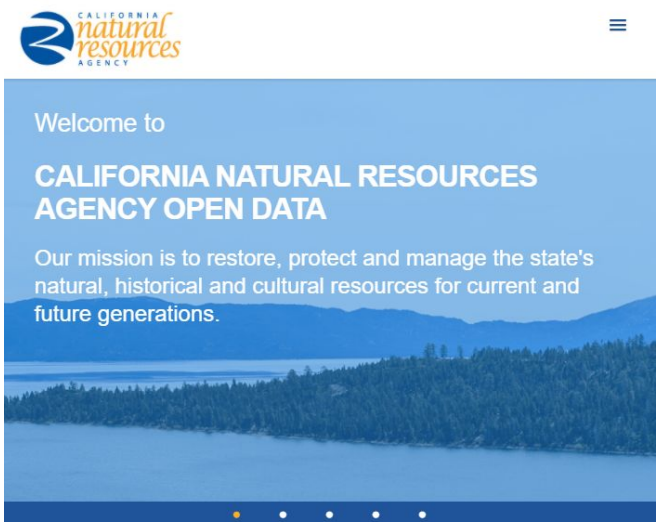
Data Openness

Open Data™

- Complete datasets
- Easily and permanently available in electronic format
- No unreasonable barriers to access or restrictions on re-use
- Machine readable, non-proprietary format
- Reliable, complete metadata
- **Does NOT** contain sensitive personal information

Data Openness

Open Data™



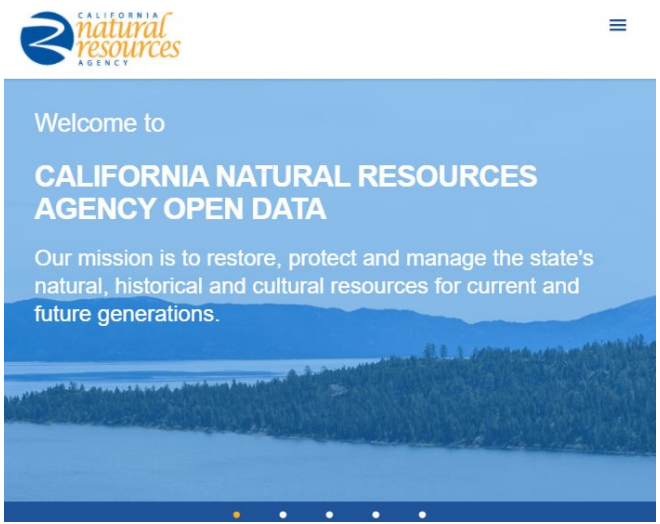
Data Openness

- ~~Complete~~ datasets
- Easily and permanently available in electronic format
- No unreasonable barriers to access or restrictions on re-use
- Machine readable, non-proprietary format
- ~~Reliable, complete metadata~~
- **Does NOT** contain sensitive personal information

"Open Data"

Data Openness

“Open Data”



Data Openness

- ~~Complete~~ datasets
- ~~Easily and permanently~~ available in electronic format
- ~~No unreasonable barriers to access or restrictions on re-use~~
- ~~Machine readable, non-proprietary format~~
- ~~Reliable, complete metadata~~
- **Does NOT** contain sensitive personal information

Public data

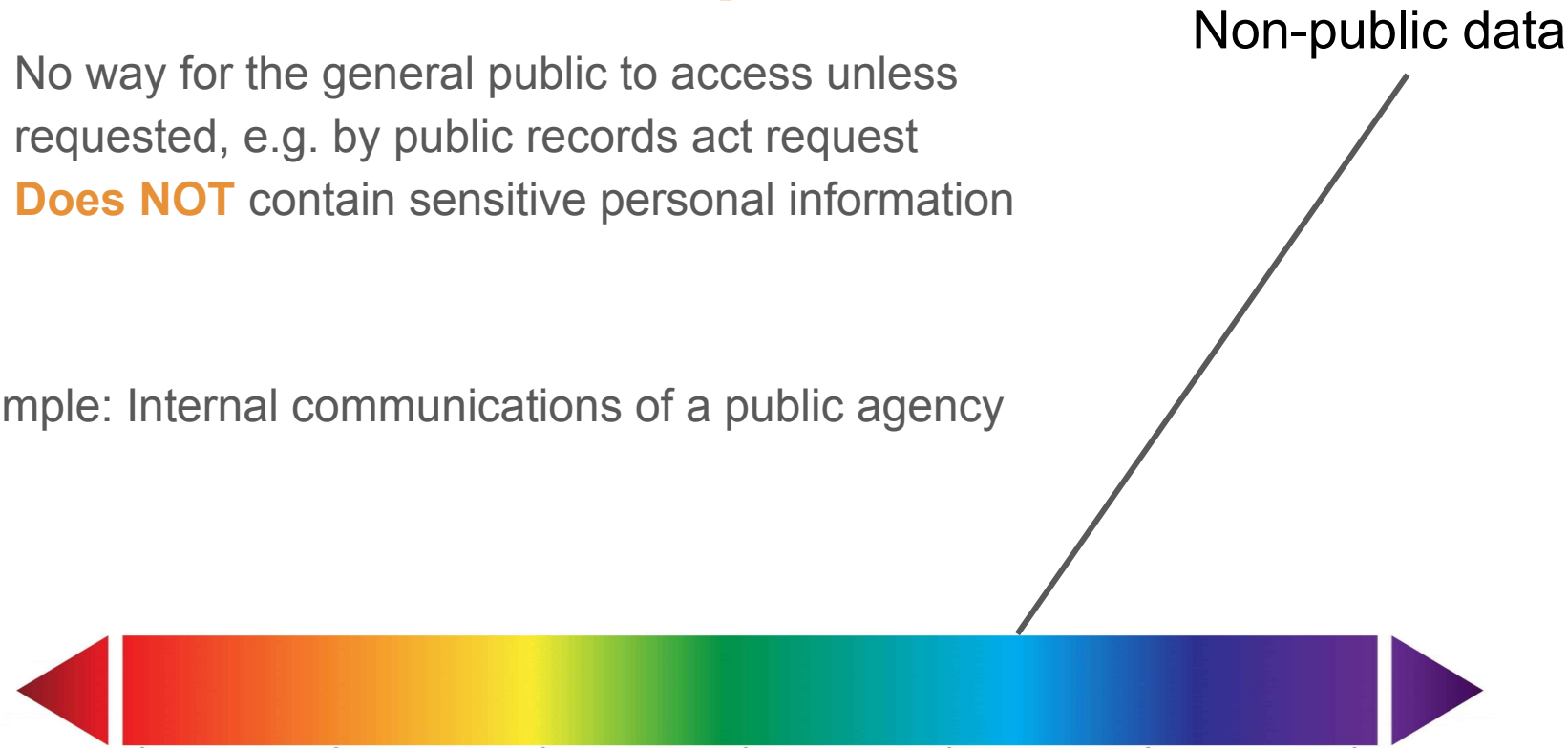
Data Openness

Public data

Data Openness

- No way for the general public to access unless requested, e.g. by public records act request
- **Does NOT** contain sensitive personal information

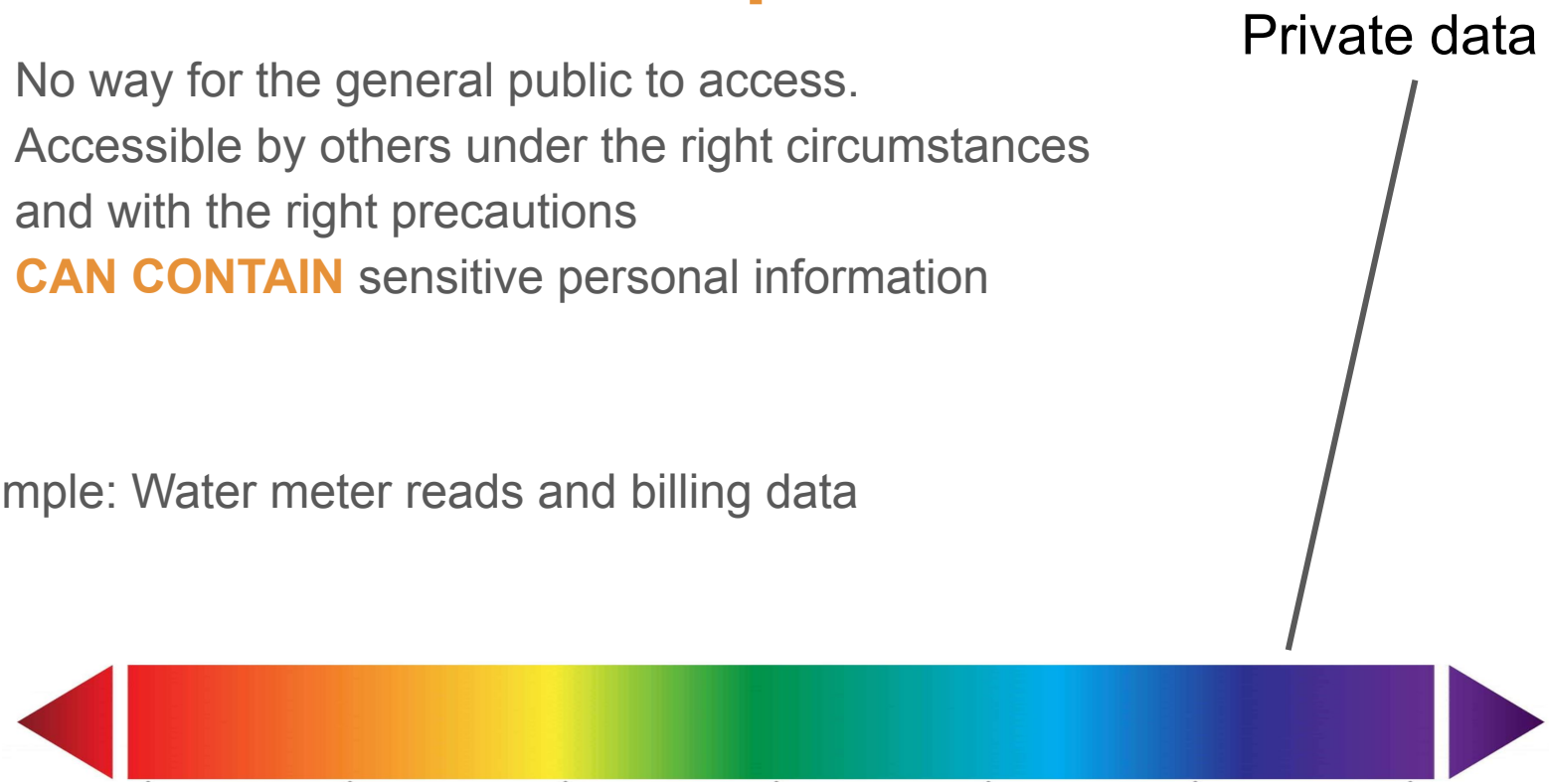
Example: Internal communications of a public agency



Data Openness

- No way for the general public to access.
- Accessible by others under the right circumstances and with the right precautions
- **CAN CONTAIN** sensitive personal information

Example: Water meter reads and billing data

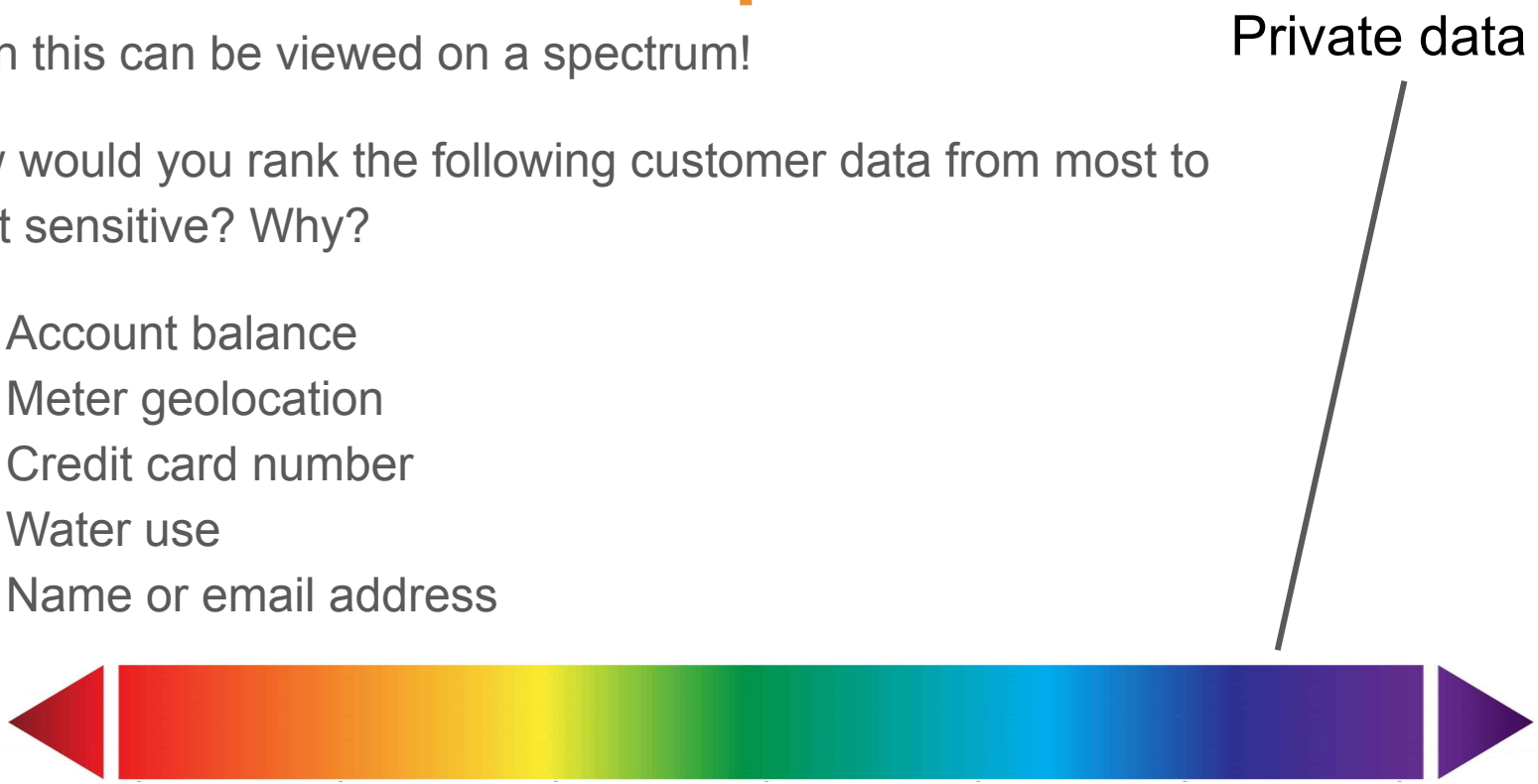


Data Openness

Even this can be viewed on a spectrum!

How would you rank the following customer data from most to least sensitive? Why?

- Account balance
- Meter geolocation
- Credit card number
- Water use
- Name or email address



Data Openness

Security-critical data

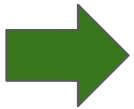
- No way for the general public to access.
- Accessible by others only when absolutely necessary and under the highest security precautions
- **CAN CONTAIN** sensitive personal information or information that threatens public safety

Example: Data that could allow access to or manipulation of critical infrastructure

Data Access

- What is an API?
- **Application Programming Interface**
- Basically a way for external programmers to interact with a software system
- Common example - retrieving data from a REST API

https://webservices.mwdsc.org/wins/Public/api/MeterInterval/OC-81/2019-10-03/2019-10-03?type=json



[{"MeterDate":"2019-10-03T00:00:00","MeterID":"OC-81","Flow":1.92,"StartDate":"1995-07-01T00:00:00","EndDate":"2045-12-31T00:00:00","BillCustID":"OC","IntervalNum":1.0,"EndMeterReading":26778939.0,"Volume":0.041322,"ProcessedFlag":"2"},{"MeterDate":"2019-10-03T00:00:00","MeterID":"OC-81","Flow":1.91,"StartDate":"1995-07-01T00:00:00","EndDate":"2045-12-31T00:00:00","BillCustID":"OC","IntervalNum":2.0,"EndMeterReading":26778956.0,"Volume":0.039027,"ProcessedFlag":"2"},{"MeterDate":"2019-10-03T00:00:00","MeterID":"OC-81","Flow":1.91,"StartDate":"1995-07-01T00:00:00","EndDate":"2045-12-31T00:00:00","BillCustID":"OC","IntervalNum":3.0,"EndMeterReading":26778973.0,"Volume":0.039027,"ProcessedFlag":"2"},...

Data About Data

Who took this photo?

When was this photo taken?

What spectral bands were collected?

Can you use this photo?

How was the photo processed?

How frequently will it be revised?

Metadata

Who took this photo? **(USDA)**

When was this photo taken? **(2018)**

What spectral bands were collected? **(RGB + NIR)**

Can you use this photo? **(Yes)**

How was the photo processed? **(Color-infrared)**

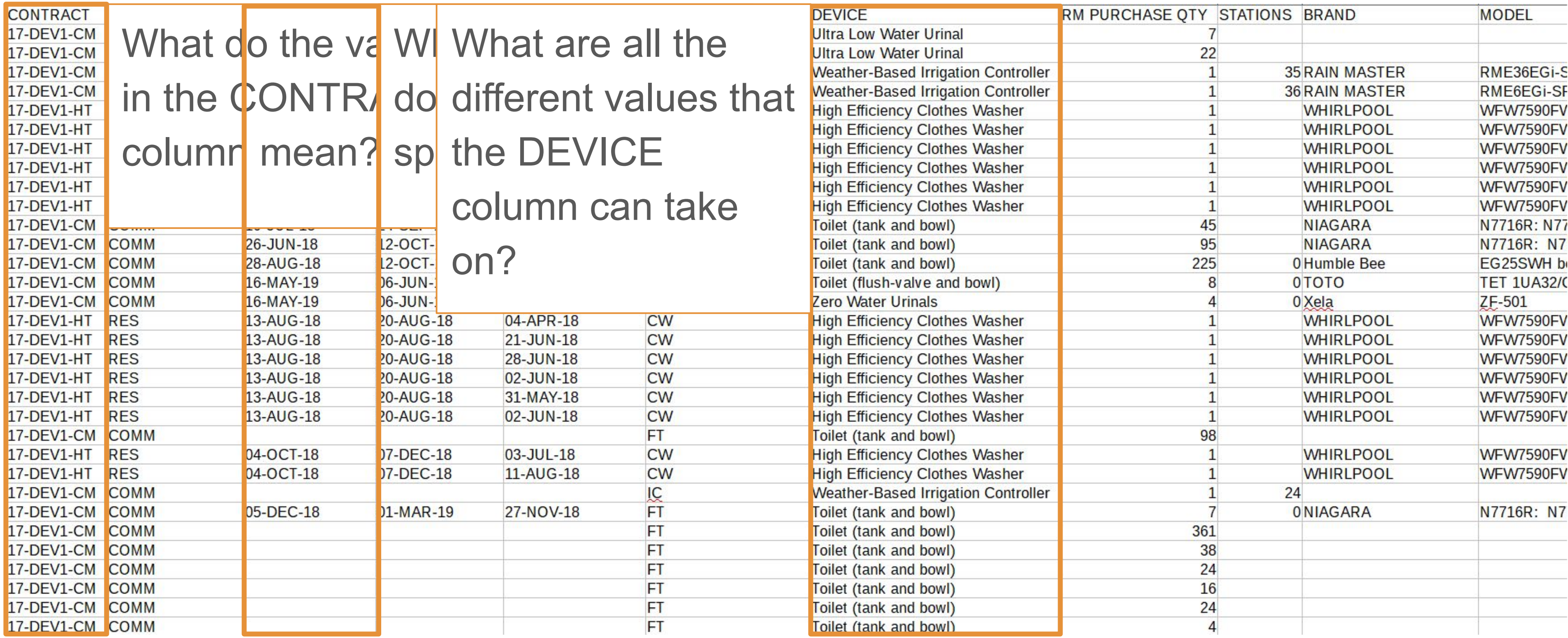
How frequently will it be revised? **(Every 2 years)**

Metadata

Good metadata...

- Makes it easy to share your data with a consultant
- Enables someone else to take over a data set you maintain
- Helps you stay sane when you revisit something years later
- Reduces the chance that data is misinterpreted

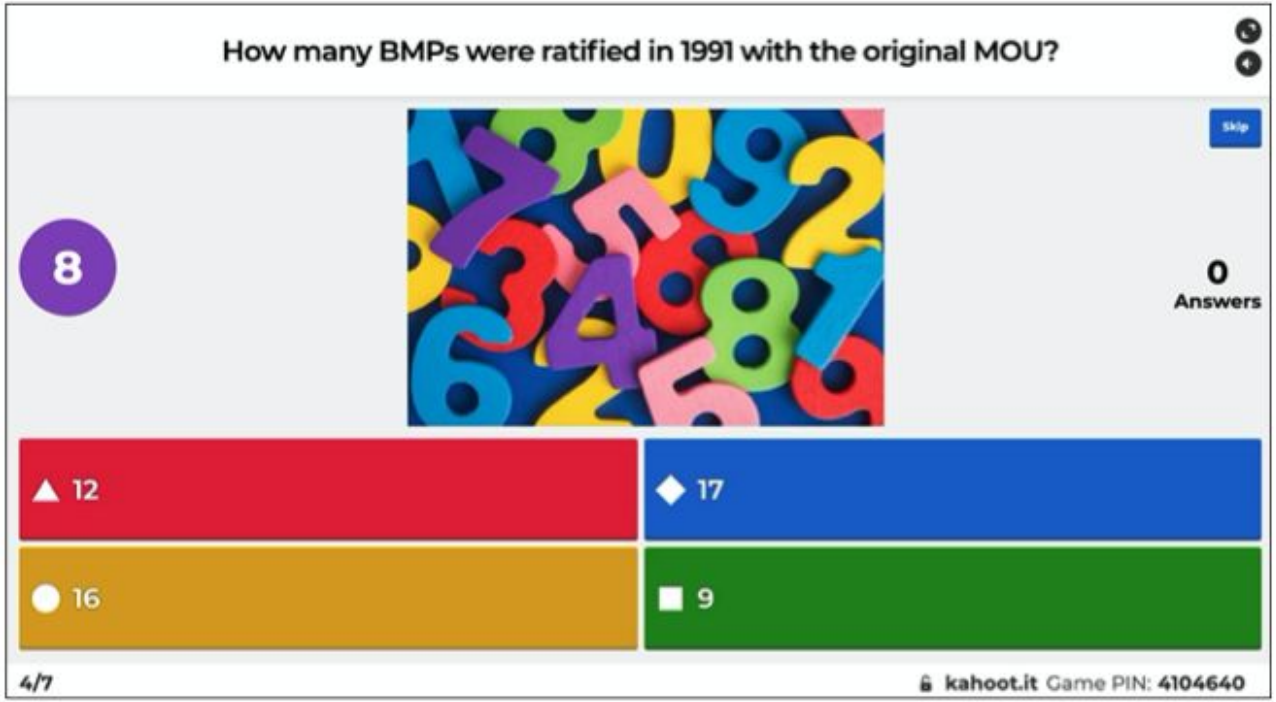
Metadata Example



Activity:
Data Terms Jeopardy

Introducing CalWEP's newest virtual addition…..

What you see on the webinar screen



What you see on your phone screen

Kahoot!

In your phone's browser open kahoot.it

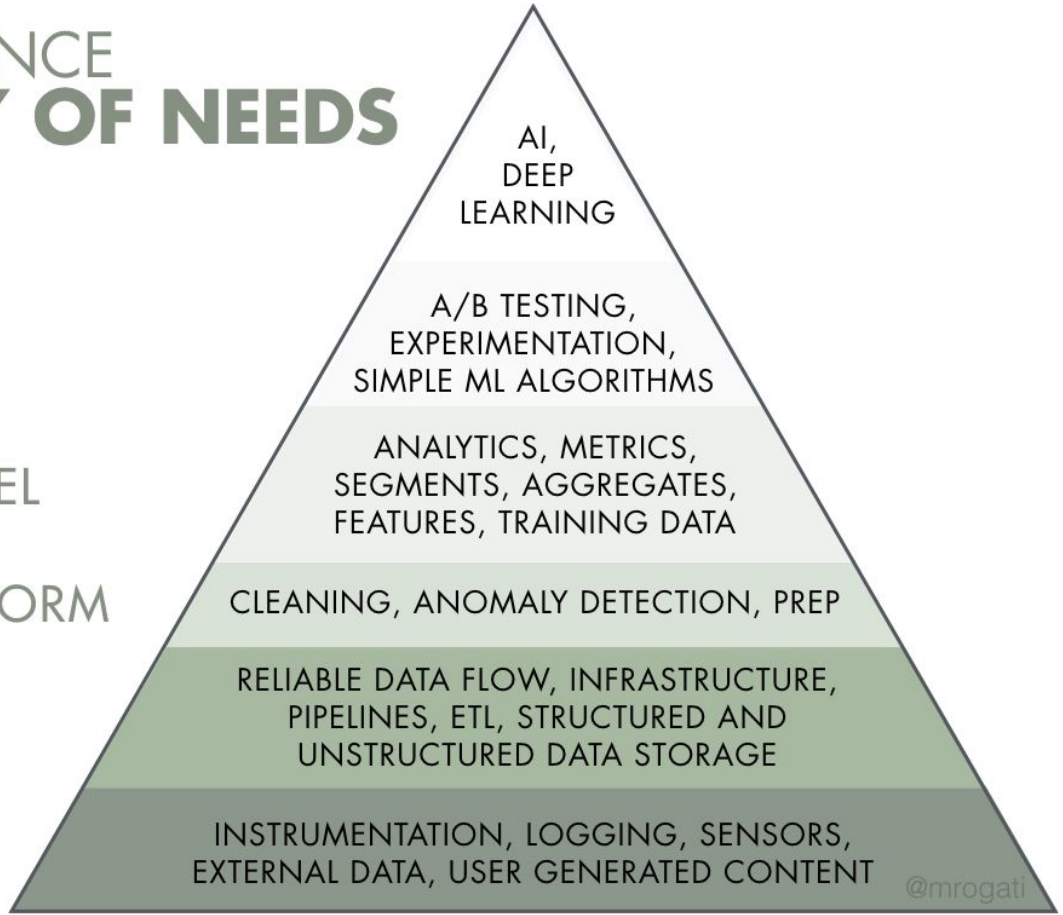
Once we launch the game you will see a pin and be asked to enter your name.

Then….we play.

(A few ways) data can be used

Check assumptions
Tell stories
Make (support) decisions
Inform questions
Enable new approaches

THE DATA SCIENCE **HIERARCHY OF NEEDS**



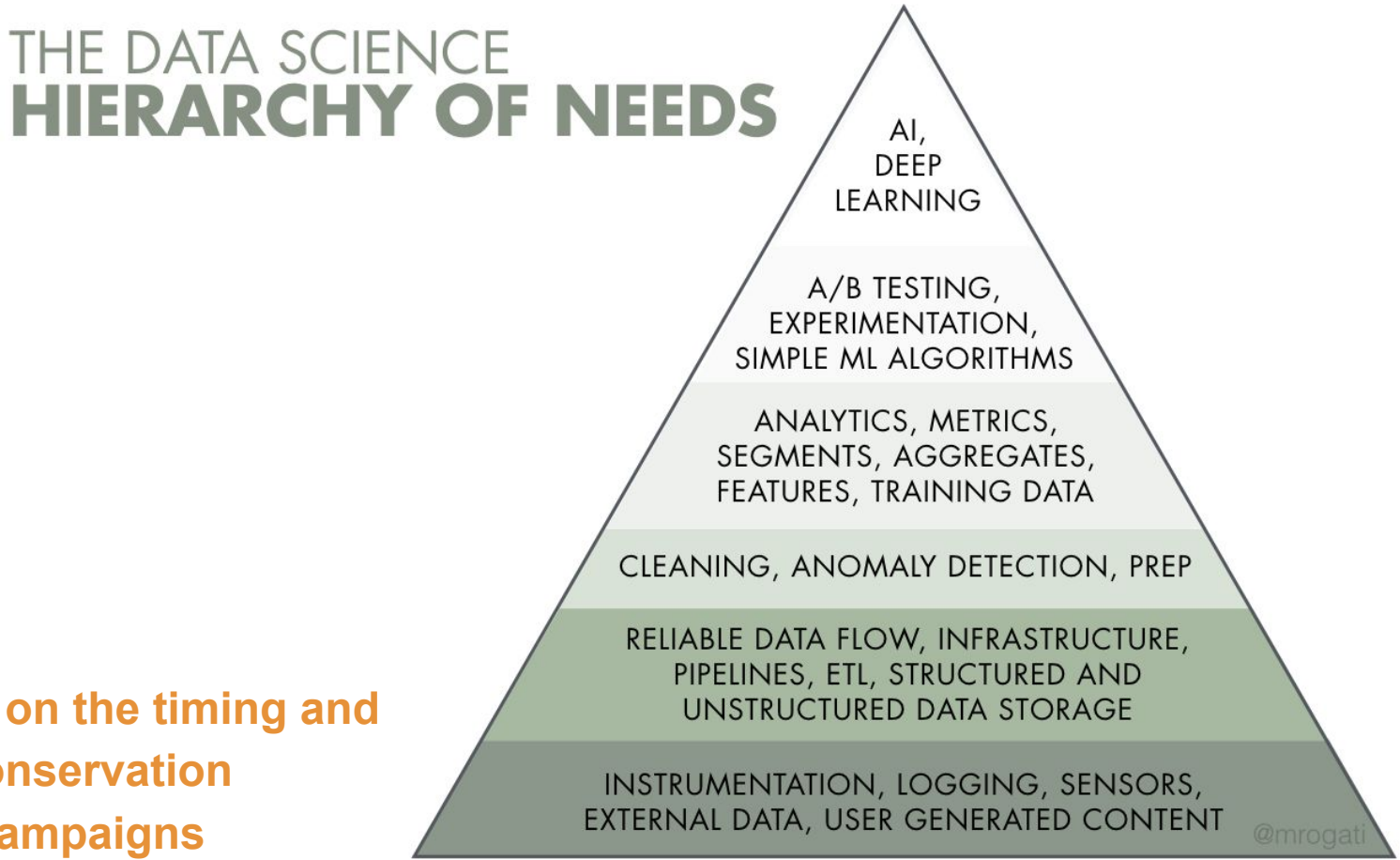
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

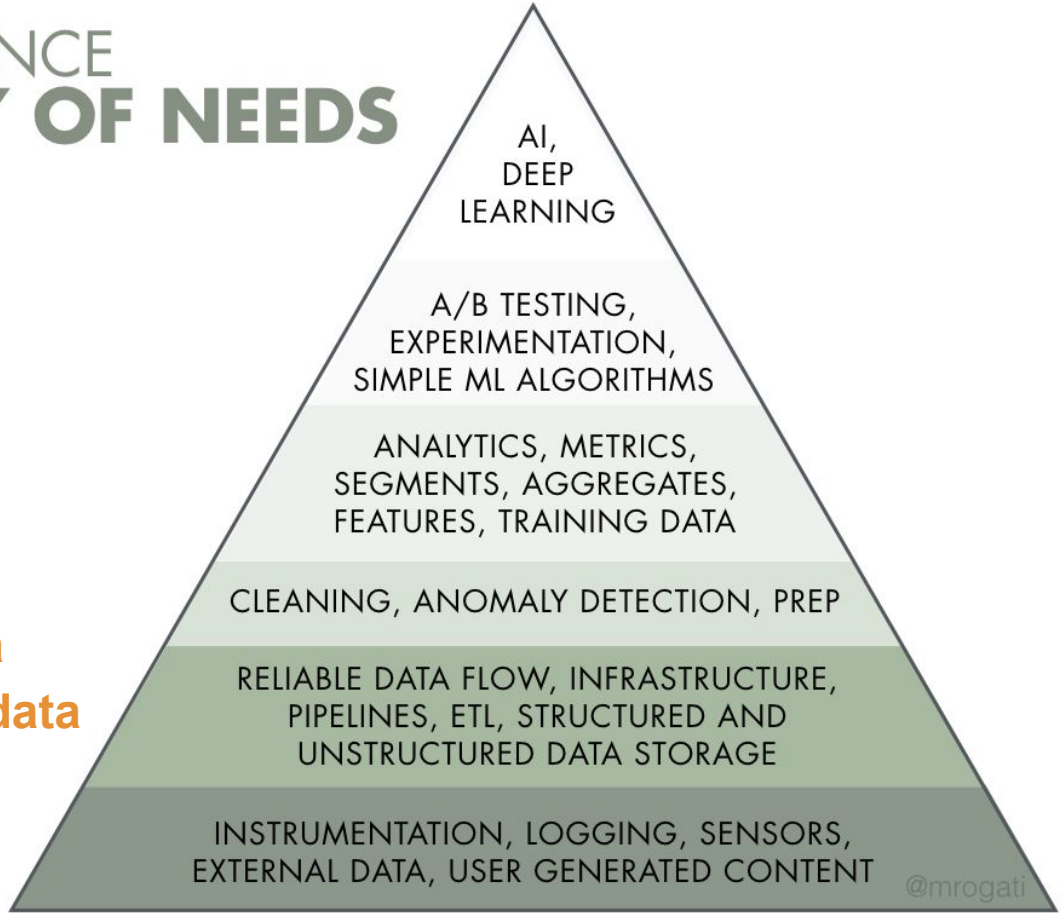
MOVE/STORE

COLLECT



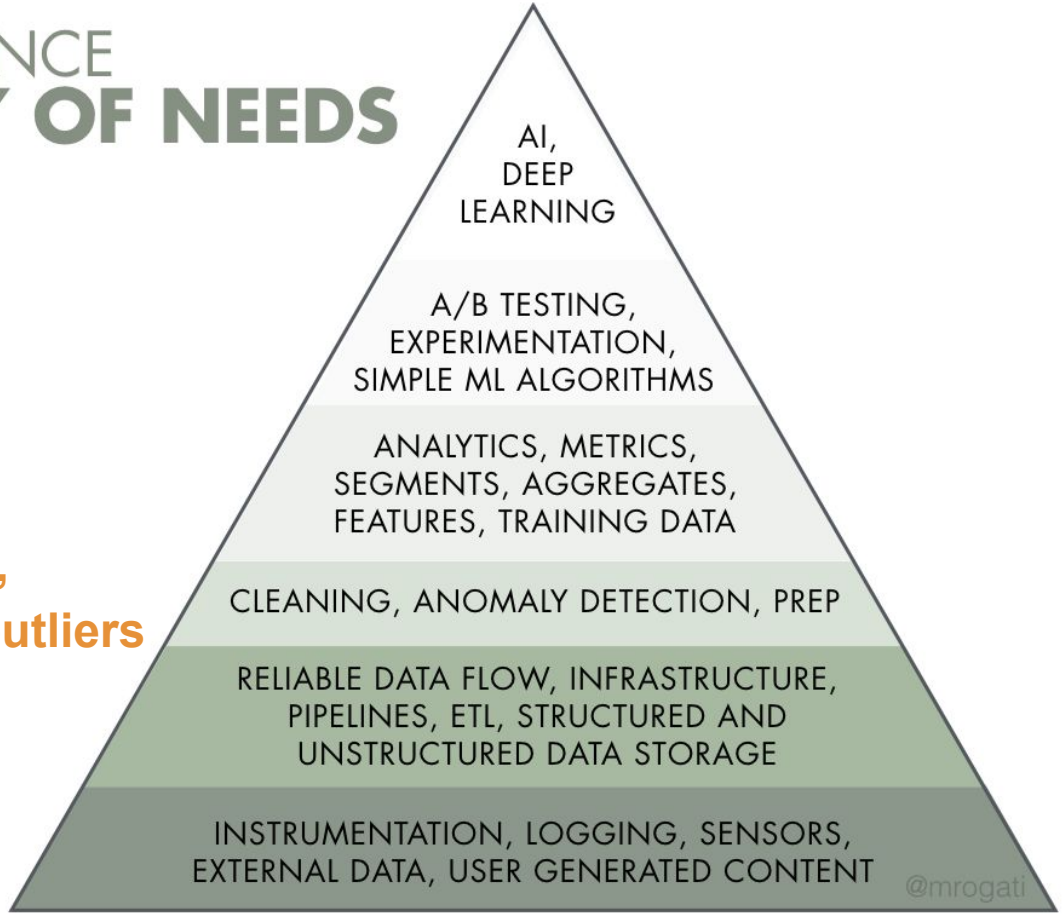
Gather data on the timing and details of conservation marketing campaigns

THE DATA SCIENCE HIERARCHY OF NEEDS



Put systems into place so data makes it to reliable long-term data storage (e.g. a database)

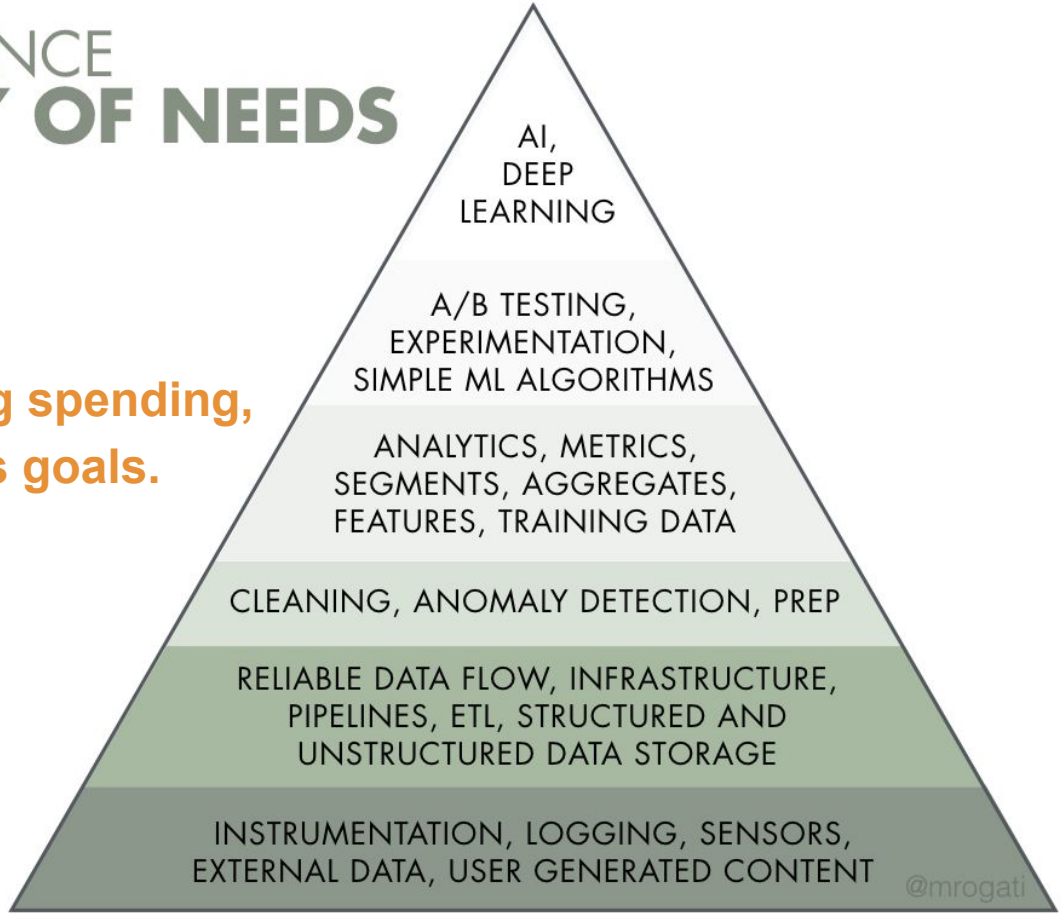
THE DATA SCIENCE HIERARCHY OF NEEDS



Remove or fill in missing fields, standardize values, check for outliers

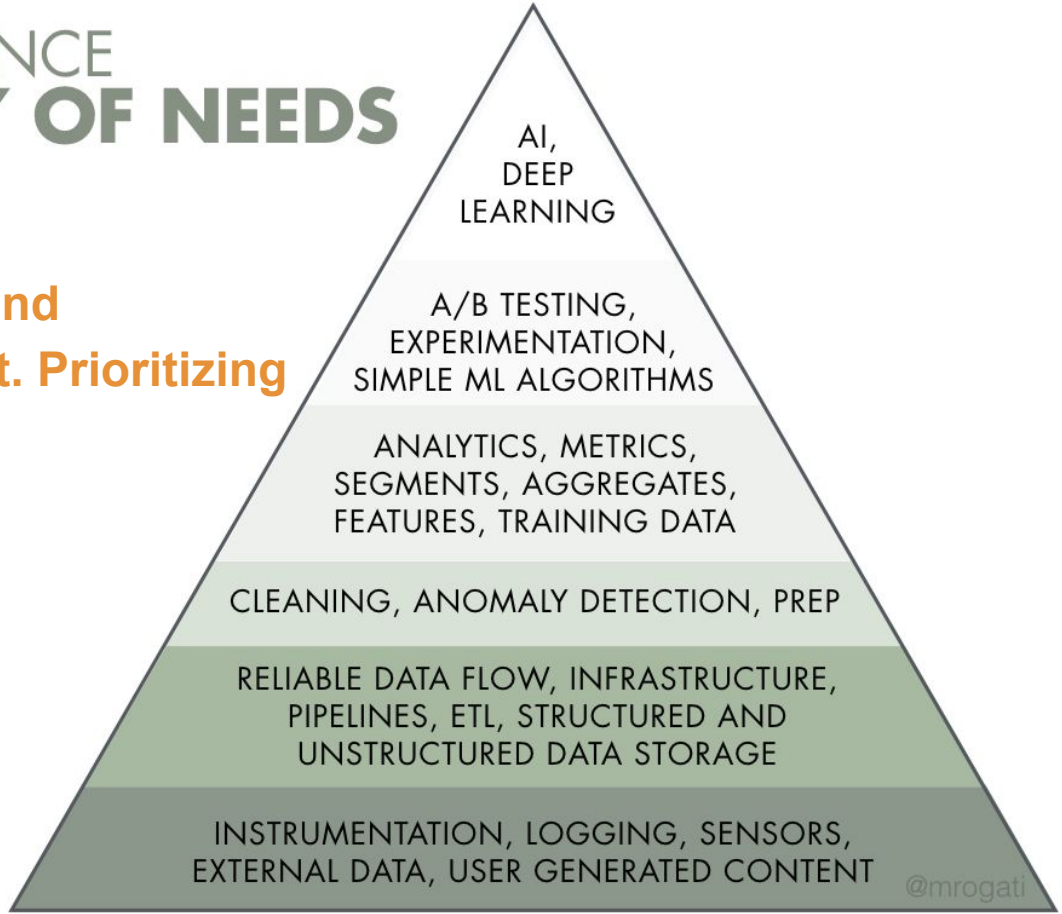
THE DATA SCIENCE HIERARCHY OF NEEDS

Dashboards or reports tracking spending, engagement, progress towards goals.



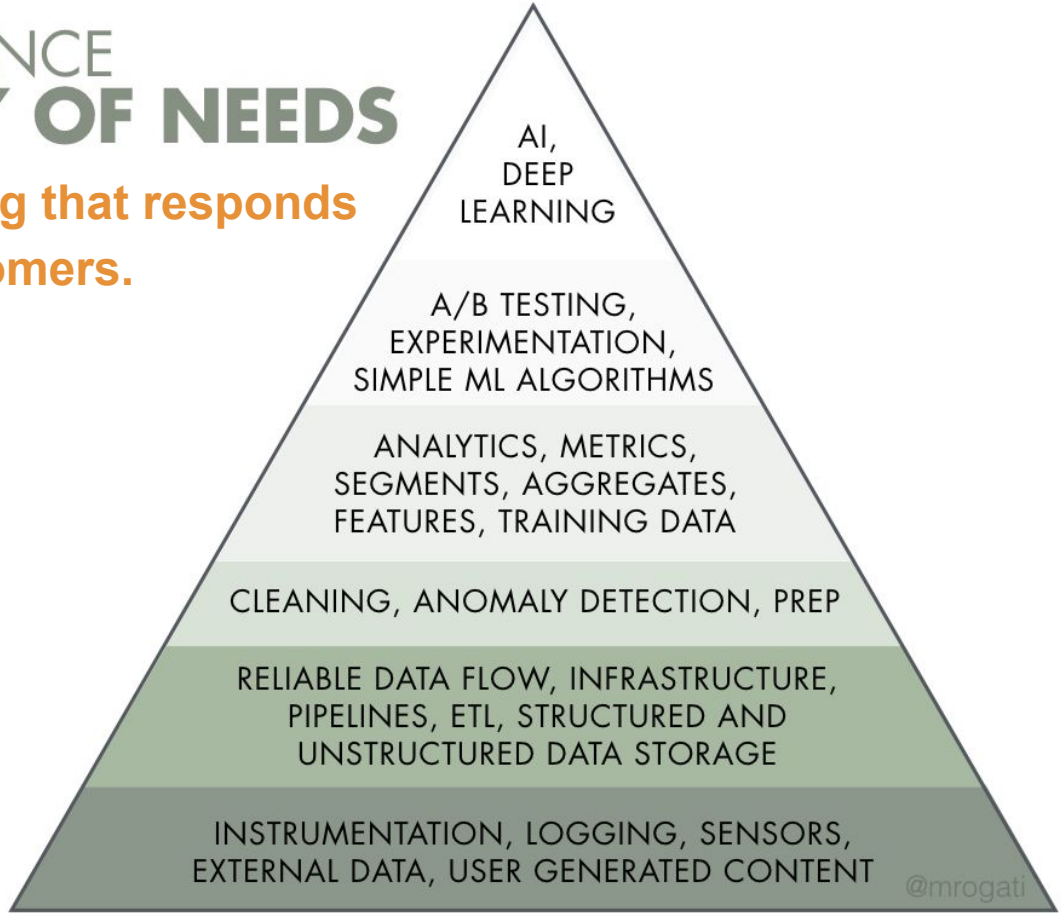
THE DATA SCIENCE HIERARCHY OF NEEDS

Testing of different messages and approaches to maximize impact. Prioritizing who to reach out to.



THE DATA SCIENCE HIERARCHY OF NEEDS

Dynamic customized messaging that responds to the needs of individual customers.



Two Purposes for Data

- Data is often originally collected for **administrative** purposes.
  - Billing customers
  - Providing a rebate check
- Administrative data is often repurposed for **analysis** after the fact.
  - Finding inefficient customers
  - Evaluating program effectiveness
- If analysis is planned for from the beginning, it usually goes more smoothly!
  - Know what you plan analyze
  - Collect the right data
  - Keep good metadata

Activity:
Data Storytelling

Tell the story of this data set:

[Purple, Beach, Button, Coffee, Experience]

Data is not neutral

Humans make choices about:

- **Data collection methods**
- **Defining terms**
- **Analyzing/applying data**
- **Visualizing data**

Minimize bias when possible, identify biases that cannot be removed, and be clear about assumptions and methodology that may affect reading of data

How many ways could you answer the question:

“How much water did your agency use last year?”

Production?
Metered?
Sales?
Calendar or fiscal year?
Pass-through?
Potable?...

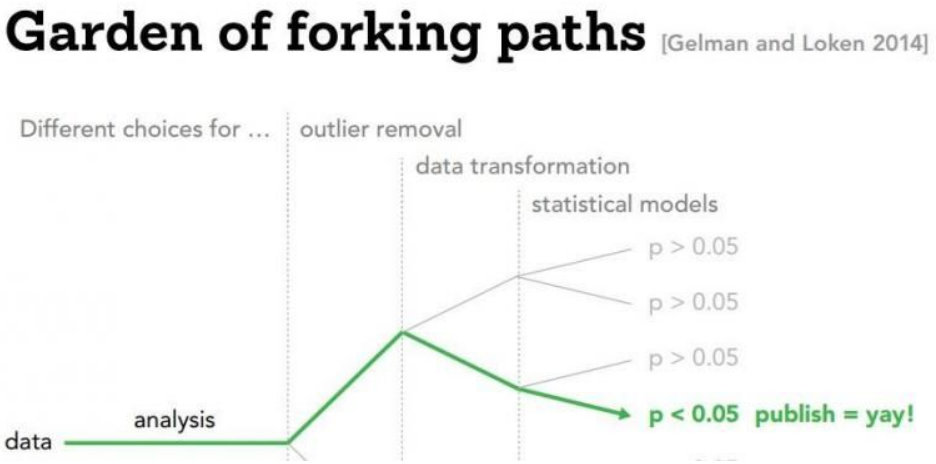
Pitfalls to watch out for

- Data isn’t what you expected
- Not sure what the data means or how/why data was collected
- Data taken out of context

Pitfalls to watch out for

“How is the sausage made?”

1. Data collection
2. Data interpretation
3. Data transformation / Analysis
4. Presentation of results
5. Interpretation of results



Applied Example:
LAM Data Assessment

Goals of LAM Assessment

1. Quality control the Landscape Area Measurement (LAM) data that DWR has provided to agencies

2. Estimate residential portions of water use objective to quantify the impact for the agency

Where to start?

Let’s break down the formula to see what data we need.

Residential Indoor Residential Outdoor

(GPCD * Pop) + (ETAF * ET0 * LA * C)

Indoor standard | Population | Outdoor standard | Reference ET | Landscape Area | Scaling Factor

Standards
Data
Constant

Zooming in on Landscape Area

DWR's LAM data is derived from many different data sources with different considerations

- Imagery
- Classification Algorithm
- Agency Boundary
- Parcel Polygons

$$(\underbrace{GPCD}_{\text{Indoor standard}} * \underbrace{Pop}_{\text{Population}}) + (\underbrace{ETAF}_{\text{Outdoor standard}} * \underbrace{ET0}_{\text{Reference ET}} * \underbrace{LA}_{\text{Landscape Area}} * \underbrace{C}_{\text{Scaling Factor}})$$

Zooming in on Landscape Area

DWR's LAM data is derived from many different data sources with different considerations

- **Imagery**
- Classification Algorithm
- Agency Boundary
- Parcel Polygons

Raster data, 4-band imagery, year, time of year, resolution, errors in images?

$$(\underbrace{GPCD}_{\text{Indoor standard}} * \underbrace{Pop}_{\text{Population}}) + (\underbrace{ETAF}_{\text{Outdoor standard}} * \underbrace{ET0}_{\text{Reference ET}} * \underbrace{LA}_{\text{Landscape Area}} * \underbrace{C}_{\text{Scaling Factor}})$$

Zooming in on Landscape Area

DWR's LAM data is derived from many different data sources with different considerations

- Imagery
- **Classification Algorithm**
- Agency Boundary
- Parcel Polygons

Preprocessing steps, type of algorithm, accuracy, format of outputs

$$(GPCD * Pop) + (ETAF * ET0 * LA * C)$$

GPCD: Indoor standard; Pop: Population; ETAF: Outdoor standard; ET0: Reference ET; LA: Landscape Area; C: Scaling Factor

Zooming in on Landscape Area

DWR's LAM data is derived from many different data sources with different considerations

- Imagery
- Classification Algorithm
- **Agency Boundary**
- Parcel Polygons

Does the boundary accurately represent the area where you provide water?

$$(GPCD * Pop) + (ETAF * ET0 * LA * C)$$

GPCD: Indoor standard; Pop: Population; ETAF: Outdoor standard; ET0: Reference ET; LA: Landscape Area; C: Scaling Factor

Zooming in on Landscape Area

DWR's LAM data is derived from many different data sources with different considerations

- Imagery
- Classification Algorithm
- Agency Boundary
- **Parcel Polygons**

Parcel georectification, alignment of land use codes with agency classes, overlap of parcels with irrigated areas

$$(\underset{\text{Indoor standard}}{GPCD} * \underset{\text{Population}}{Pop}) + (\underset{\text{Outdoor standard}}{ETAF} * \underset{\text{Reference ET}}{ET0} * \underset{\text{Landscape Area}}{LA} * \underset{\text{Scaling Factor}}{C})$$

Zooming in on Landscape Area

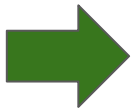
- Only way to check many of these things is to open up the files and look!
- Shapefiles opened in ArcGIS, QGIS or similar
- Compare parcel locations against residential meter locations

$$(\underbrace{GPCD}_{\text{Indoor standard}} * \underbrace{Pop}_{\text{Population}}) + (\underbrace{ETAF}_{\text{Outdoor standard}} * \underbrace{ET0}_{\text{Reference ET}} * \underbrace{LA}_{\text{Landscape Area}} * \underbrace{C}_{\text{Scaling Factor}})$$

Zooming in on Evapotranspiration

- Where to source data on reference evapotranspiration?
- CIMIS is standard. Spatial CIMIS vs. Individual Stations.
- CIMIS has a REST API!

```
http://et.water.ca.gov/api/data?appKey=YOUR-APP-KEY&targets=2,8,127&startDate=2010-01-01&endDate=2010-01-05
```




```
{
  "Data": {
    "Providers": [
      {
        "Name": "cimis",
        "Type": "station",
        "Owner": "water.ca.gov",
        "Records": [
          {
            "Date": "2010-01-01",
            "Julian": "1",
            "Station": "2",
            "Standard": "english",
            "ZipCodes": "93624",
            "Scope": "daily",
            "DayAirTmpAvg": {
              "Value": "39",
              "Qc": " ",
              "Unit": "(F)"
            },
            "DayAirTmpMax": {
              "Value": "57.3",
              "Qc": " ",
              "Unit": "(F)"
            },
            "DayAirTmpMin": {
              "Value": "29.8",
              "Qc": " ",
              "Unit": "(F)"
            },
            ... Results Truncated ...
          }
        ]
      }
    ]
  }
}
```


$$(GPCD * Pop) + (ETAF * ET0 * LA * C)$$

Indoor standard | Population | Outdoor standard | Reference ET | Landscape Area | Scaling Factor

Zooming in on Population

- How to calculate population?
  - Number of meters * Household Size
  - Census
  - American Community Survey

- Seasonal population?

$$(GPCD * Pop) + (ETAF * ET0 * LA * C)$$

GPCD: Indoor standard; Pop: Population; ETAF: Outdoor standard; ET0: Reference ET; LA: Landscape Area; C: Scaling Factor

Today we talked about:

- The Universe of Water Agency Data
- Data Types and Formats
- Accessing and Sharing Data
- Metadata
- Data Terms Kahoot!
- Potential and Pitfalls
- An Applied Example

Resources

- Data Terms
  - https://www.springboard.com/blog/data-science-terms/
  - https://www.dataquest.io/blog/data-science-glossary/
- Data needs pyramid
  - https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007
-

Upcoming Events

More info: CalWEP.org

CA Water Data Summit

August 19-20, 2021

More info: CaWaterDataSummit.org

Questions?