

CMPU 187 · Introduction to Digital Humanities

# Data

21 April 2021



What's data?

“The most inclusive summary is to say that data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.”

Christine Borgman, *Big Data, Little Data, No Data*

What's data good for?  
What's data bad for?

Working with data

NJShipwrecks											
Sheet 1											
Table data was imported. <a href="#">Adjust Settings</a>											
SHIP'S NAME	AKA	SHIP'S OWNER	VESSEL TYPE	YEAR BUILT	WHERE BUILT	DATE LOST	YEAR	MNTH	DAY	LOCATION LOST	LATITUDE LO
74 H	93015		Scow			7/28/1916	1916	7	28	Sandy Hook SE 5 mi	
A B Crosby			Schooner	1884		3/12/1888	1888	3	12	Atlantic Highlands	
A B Sherman			Schooner	1883		12/8/1886	1886	12	8	Five Fathoms Bank Light	
A B Thompson			Yacht			7/18/1874	1874	7	18	Absecon	
A C Austin			Schooner			2/1868	1868	2		Sandy Hook Romer Shoal	
A C Wescoat (1989)	\	A C Wescoat Co	Deck barge			10/11/1989	1989	10	11	Atlantic City Artificial Reef	39-15-540 N
A C Wescoat (2006)			Deck barge			7/26/2006	2006	7	25	Ocean City Artificial Reef	39-09-819 N
A D Scull			Schooner	1864		4/18/1880	1880	4	18	Townsend's Inlet	
A E Douglass			Schooner	1855		6/4/1862	1862	6	4	Barnegat N 8 mi	
A F Baillie			Schooner	1872		11/17/1875	1875	11	17	Carters Bar, VA	
A F Crockett			Schooner	1881	Bath, ME	5/21/1894	1894	5	21	Harvey Cedars NE .5 mi	
A Field	A D Field		Schooner			2/10/1852	1852	2	10	Deal Beach	
A G Ropes		Luckenbach SS Co	Schooner - Barge	1884	Bath, ME	12/26/1913	1913	12	26	Island Beach	
A H Bowman			Steamship			12/4/1857	1857	12	4	Deal Lake N .25 mi	
A H DuMont		Spentonbush Red Star Co	Tanker	1940	Camden, NJ	6/10/1986	1986	6	10	Garden State North Artificial Reef	39-37-727 N
A H Edward			Schooner			1820	1820			Nine Mile Beach	
A H Hurlburt			Schooner	1871	Middletown, CT	6/26/1884	1884	6	26	Cape May LSS SW 4 mi	
A J Bentley		Bentley & Gildersleeve	Schooner	1865	Hartford, CT	5/27/1873	1873	5	27	Barnegat	
A J Hoole			Schooner			12/19/1902	1902	12	19	Spermaceti Cove LSS S 1 mi	
A L & M Townsend	George White		Schooner	1871	Barnegat, NJ	12/30/1888	1888	12	30	Little Egg LSS S 2 mi	
A L Lee (1897)			Schooner	1884	Mays Landing, NJ	3/25/1897	1897	3	25	Ocean City LSS E .5 mi	
A L Lee (1898)			Schooner	1884	Mays Landing, NJ	8/26/1898	1898	8	26	Absecon bar	
A L Lee (1902)			Schooner	1884	Mays Landing, NJ	12/10/1902	1902	12	10	Absecon bar	
A Lincoln			Schooner	1865	Essex, MA	10/12/1891	1891	10	12	Barnegat	

Data: <https://bit.ly/3xfMAEQ>

# CSV: Comma-separated value

Fruit,Price

Apple,\$0.50

Banana,\$10



# CSV: Comma-separated value

Song, Lead, Year  
Help!, Lennon, 1965



# CSV: Comma-separated value

Song,Lead,Year

Help!,Lennon,1965

Can't Buy Me Love,McCartney,1969

# CSV: Comma-separated value

Song,Lead,Year

Help!,Lennon,1965

Can't Buy Me Love,McCartney,1969

Here, There and Everywhere,McCartney,1966

# CSV: Comma-separated value

Song,Lead,Year

Help!,Lennon,1965

Can't Buy Me Love,McCartney,1969

Here, There and Everywhere,McCartney,1966



*Uh oh*

# CSV: Comma-separated value

Song, Lead, Year

Help!, Lennon, 1965

Can't Buy Me Love, McCartney, 1969

Here, There and Everywhere, McCartney, 1966

Uh oh



*The comma llama strikes again!*

<https://www.youtube.com/watch?v=fEcKltDBQjI>

# CSV: Comma-separated value

"Song", "Lead", "Year"

"Help!", "Lennon", "1965"

"Can't Buy Me Love", "McCartney", "1969"

"Here, There and Everywhere", "McCartney", "1966"

*Better, but problems can still occur if we move on from the Beatles*

# CSV: Comma-separated value

"Song", "Lead", "Year"

"Help!", "Lennon", "1965"

"Can't Buy Me Love", "McCartney", "1969"

"Here, There and Everywhere", "McCartney", "1966"

""Heroes"" , "Bowie", "1977"

*Bowie scares us with his quotation marks.*

# CSV: Comma-separated value

"Song","Lead","Year"

"Help!","Lennon","1965"

"Can't Buy Me Love","McCartney","1969"

"Here, There and Everywhere","McCartney","1966"

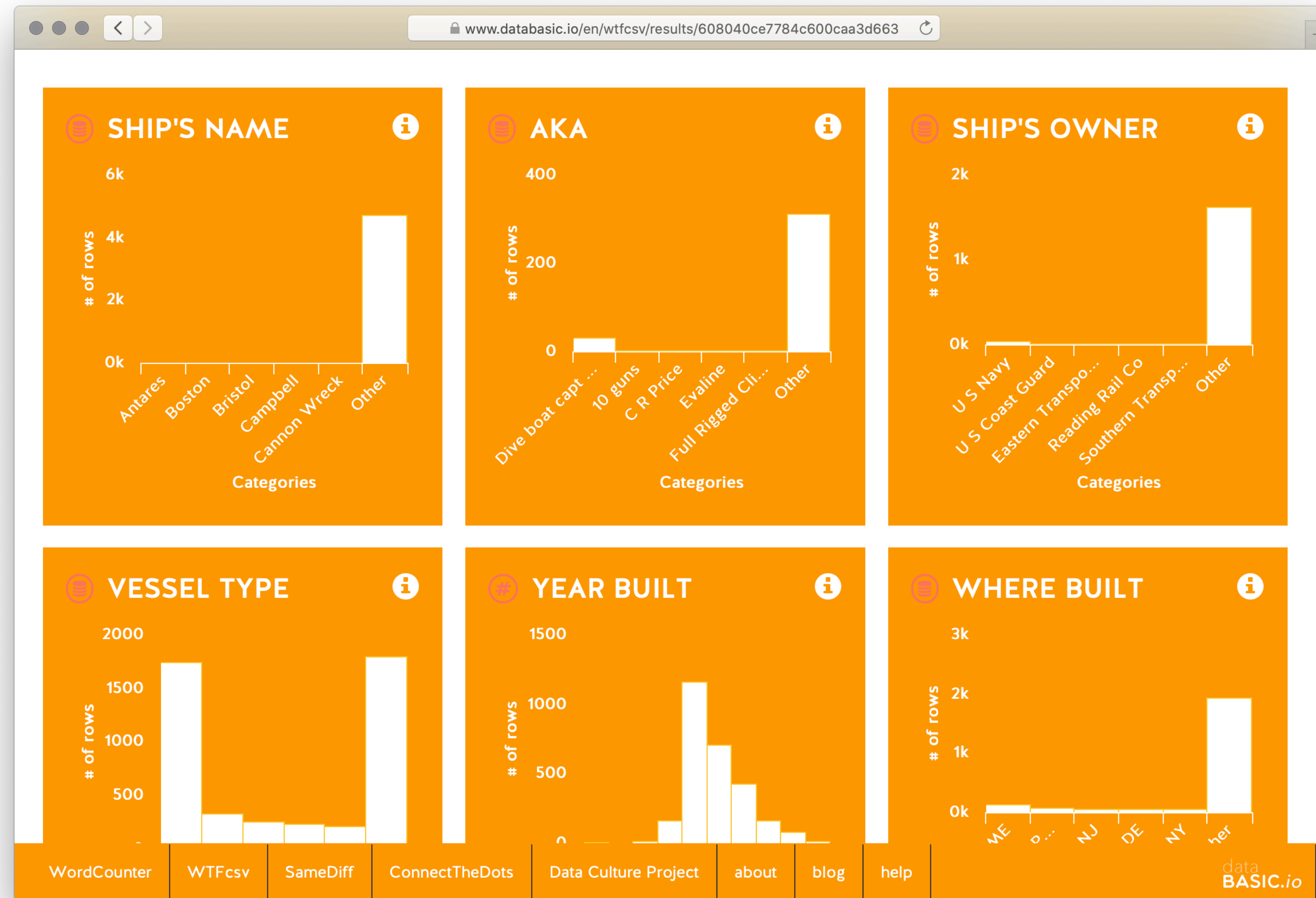
"\"Heroes\"","Bowie","1977"

*Sorted.*

# TSV: Better living through tabs

Song	Lead	Year
Help!	Lennon	1965
Can't Buy Me Love	McCartney	1969
Here, There and Everywhere	McCartney	1966
"Heroes"	Bowie	1977





*[databasic.io/en/wtfcsv](http://databasic.io/en/wtfcsv)*

What are some questions you could explore with this data set?

# XML gives us nesting

```
<dvdCollection>
  <dvd>
    <id>1</id>
    <title>Raiders of the Lost Ark</title>
    <release-year>1981</release-year>
    <director>Steven Spielberg</director>
    <actors>
      <actor>Harrison Ford</actor>
      <actor>Karen Allen</actor>
      <actor>John Rhys-Davies</actor>
    </actors>
  </dvd>
  ...
</dvdCollection>
```



# JSON gives us nesting with less crying

```
{
  "dvds": [
    "1": {
      "title": "Raiders of the Lost Ark",
      "release year": 1981,
      "director": "Steven Spielberg",
      "actors": [
        "Harrison Ford",
        "Karen Allen",
        "John Rhys-Davies"
      ]
    },
    ...
  ]
}
```

