

Week 2 | Speed data-ing

Data sources

- The Internet/Google
- The government (local, national, agencies)
 - International organizations
- Think tanks and research organizations
 - NGOs and advocacy organizations
 - Private industry groups
 - Scientific research (ex 1 ex 2)
 - Collect it yourself
 - FOIA/FOIL

Examples

Opening up data:

- 2012 Olympic athletes – The Guardian
 - Dogs of NYC – WNYC
- The Military's Surplus Gear – New York Times (bonus link)
 - The Counted – The Guardian

Linking to data:

- Why movies are terrible in the winter – FiveThirtyEight

Common data formats

Human v machine readable

PDF

name	age_range	gender	borough
Student 1	36-40	Female	Brooklyn
Student 2	18-23	Female	Manhattan
Student 3	18-23	Female	Queens
Student 4	24-29	Male	Brooklyn
Student 5	24-29	Male	Brooklyn
Student 6	24-29	Male	Queens
Student 7	24-29	Male	Brooklyn
Student 8	18-23	Female	Manhattan
Student 9	24-29	Female	New Jersey
Student 10	24-29	Female	Manhattan
Student 11	18-23	Female	Queens
Student 12	18-23	Female	Brooklyn
Student 13	24-29	Female	Brooklyn
Student 14	18-23	Male	Brooklyn
Student 15	24-29	Male	Manhattan



XLS (Excel)



Fundamentals of data and interactive journalism survey (Responses) ☆ 📁

File Edit View Insert Format Data Tools Form Add-ons Help All changes saved in Drive

🖨️ ↶ ↷ 📄 \$ % .0 ↓ .00 ↘ 123 ▾ Arial ▾ 10 ▾ **B** *I* ~~U~~ A ▾ 🎨 🏠 📏

	A	B	C	D	E
1	name	age_range	gender	borough	
2	Student 1	36-40	Female	Brooklyn	
3	Student 2	18-23	Female	Manhattan	
4	Student 3	18-23	Female	Queens	
5	Student 4	24-29	Male	Brooklyn	
6	Student 5	24-29	Male	Brooklyn	
7	Student 6	24-29	Male	Queens	
8	Student 7	24-29	Male	Brooklyn	
9	Student 8	18-23	Female	Manhattan	
10	Student 9	24-29	Female	New Jersey	
11	Student 10	24-29	Female	Manhattan	
12	Student 11	18-23	Female	Queens	
13	Student 12	18-23	Female	Brooklyn	
14	Student 13	24-29	Female	Brooklyn	
15	Student 14	18-23	Male	Brooklyn	
16	Student 15	24-29	Male	Manhattan	
17					

CSV

```
name,age_range,gender,borough  
Student 1,36-40,Female,Brooklyn  
Student 2,18-23,Female,Manhattan  
Student 3,18-23,Female,Queens  
Student 4,24-29,Male,Brooklyn  
Student 5,24-29,Male,Brooklyn  
Student 6,24-29,Male,Queens  
Student 7,24-29,Male,Brooklyn  
Student 8,18-23,Female,Manhattan  
Student 9,24-29,Female,New Jersey  
Student 10,24-29,Female,Manhattan  
Student 11,18-23,Female,Queens  
Student 12,18-23,Female,Brooklyn  
Student 13,24-29,Female,Brooklyn  
Student 14,18-23,Male,Brooklyn  
Student 15,24-29,Male,Manhattan
```

JSON

```
[  
  {  
    "name": "Student 1",  
    "age_range": "36-40",  
    "gender": "Female",  
    "borough": "Brooklyn"  
  },  
  {  
    "name": "Student 2",  
    "age_range": "18-23",  
    "gender": "Female",  
    "borough": "Manhattan"  
  },  
  {  
    "name": "Student 3",  
    "age_range": "18-23",  
    "gender": "Female",  
    "borough": "Queens"  
  },  
  {  
    "name": "Student 4",  
    "age_range": "24-29",  
    "gender": "Male",  
    "borough": "Brooklyn"  
  },  
  {  
    "name": "Student 5",  
    "age_range": "24-29",  
    "gender": "Male",  
    "borough": "Brooklyn"  
  }  
]
```


XML

```
<class>
  <student name='Student 1'>
    <age_range>36-40</age_range>
    <gender>Female</gender>
    <borough>Brooklyn</borough>
  </student>
  <student name='Student 2'>
    <age_range>18-23</age_range>
    <gender>Female</gender>
    <borough>Manhattan</borough>
  </student>
  <student name='Student 3'>
    <age_range>18-23</age_range>
    <gender>Female</gender>
    <borough>Queens</borough>
  </student>
  <student name='Student 4'>
    <age_range>24-29</age_range>
    <gender>Male</gender>
    <borough>Brooklyn</borough>
  </student>
  <student name='Student 5'>
    <age_range>24-29</age_range>
    <gender>Male</gender>
    <borough>Brooklyn</borough>
  </student>
  <student name='Student 6'>
    <age_range>24-29</age_range>
    <gender>Male</gender>
    <borough>Queens</borough>
  </student>
</class>
```

Let's get some data ...

<http://www.nyc.gov/data>

BREAK

Spreadsheet exercises

<http://bit.ly/spreadsheet-basics>

Slides made using [Reveal.js](#)