

INST728E - MODULE 3

# JSON DATA AND PARSING SOCIAL MEDIA

CODY BUNTAIN  
@CODYBUNTAIN  
[CBUNTAIN@CS.UMD.EDU](mailto:CBUNTAIN@CS.UMD.EDU)

# MODULE 3

## PARSING SOCIAL MEDIA DATA

```
{
  "contributors": null,
  "truncated": false,
  "text": "@megturney After seeing your tweets about #Ferguson I decided to look it up. The situation is horrible!",
  "in_reply_to_status_id": null,
  "id": 499922742045868000,
  "favorite_count": 0,
  "source": "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone</a>",
  "retweeted": false,
  "coordinates": {
    "type": "Point",
    "coordinates": [
      -79.11194229,
      43.07537783
    ]
  },
},
"entities": {
  "symbols": [],
  "user_mentions": [
    {
      "indices": [
        0,
        10
      ],
      "screen_name": "megturney",
      "id": 23711472,
      "name": "Meg Turney",
      "id_str": "23711472"
    }
  ],
  "hashtags": [
    {
      "indices": [
        42,
        51
      ]
    }
  ]
}
```

# WHAT IS JSON?

- Similar to XML but less structured
- More flexible than CSV files
- Text version of Python's dictionary
  - key: value pairs
- Why do we care?
  - Often used as a data exchange/storage format
  - Most API responses are now JSON

```
{
  "contributors": null,
  "truncated": false,
  "text": "@megturney After seeing your tweets about #Ferguson I decided to",
  "in_reply_to_status_id": null,
  "id": 499922742045868000,
  "favorite_count": 0,
  "source": "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'",
  "retweeted": false,
  "coordinates": {
    "type": "Point",
    "coordinates": [
      -79.11194229,
      43.07537783
    ]
  },
  "entities": {
    "symbols": [],
    "user_mentions": [
      {
        "indices": [
          0,
          10
        ],
        "screen_name": "megturney",
        "id": 23711472,
        "name": "Meg Turney",
        "id_str": "23711472"
      }
    ],
    "hashtags": [
      {
        "indices": [
          42,
          51
        ],
        "text": "Ferguson"
      }
    ]
  },
  "urls": []
},
```

# PARSING JSON

- Python has good support for JSON with the `json` package
- ```
obj = json.loads("{\"key\": \"value\"}")  
obj["key"] == "value"
```
- We'll use this library to parse Twitter data

# MODULE 3 HOMEWORK

# FILL OUT AND SUBMIT MODULE 03 NOTEBOOK

- Similar to Module 2 but with tweet data
- I.e., given a set of tweets, find most common tokens, hashtags, and users

# ADDITIONAL RESOURCES

# ADDITIONAL RESOURCES

- YouTube video on Python and JSON
- <https://youtu.be/9N6a-VLBa2I>



# JUPYTER NOTEBOOK EXAMPLE