Forritunarmálið Python

Day 7

Data

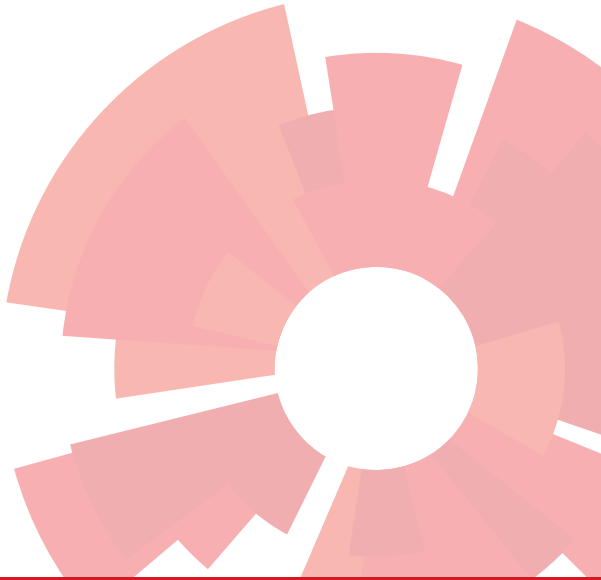
Hjalti Magnússon

4. desember 2017

Data

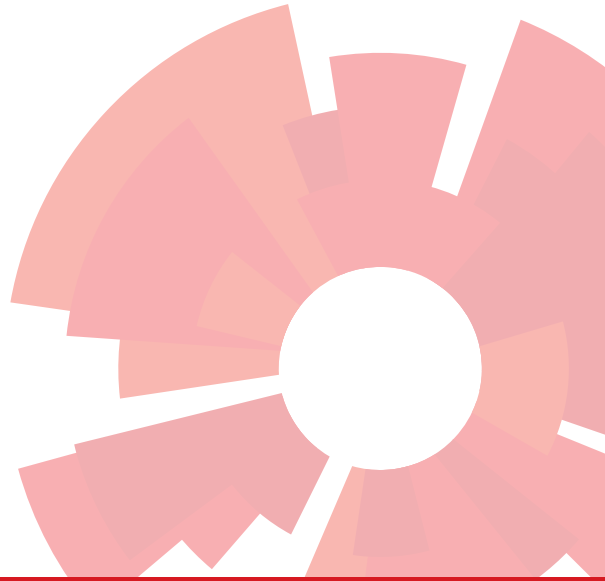
Structured vs unstructured

- We have been working mainly with unstructured data
  - No formal definition
  - Possible ambiguities
- Structured data
  - Formal representation of data
  - Well defined syntax
  - No ambiguities



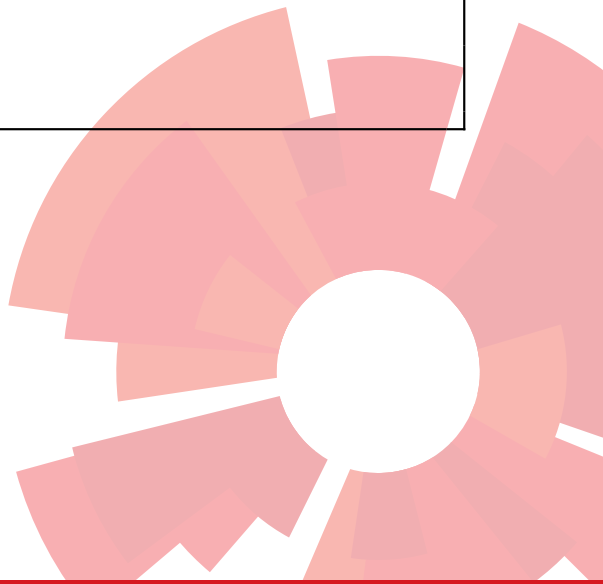
JSON

- JavaScript Object Notation
- Consists of
  - objects (dictionaries)
  - arrays (lists)
  - strings
  - numbers
  - booleans
  - null



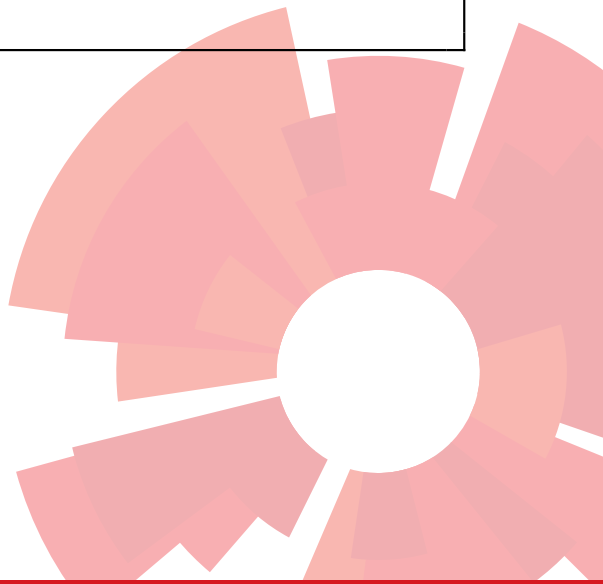
JSON

```
>>> import json
>>> js = '''{
...     "key1": 3,
...     "key2": 43.2,
...     "list": [1, 2, 3],
...     "boolean": true,
...     "test": null
... }'''
```



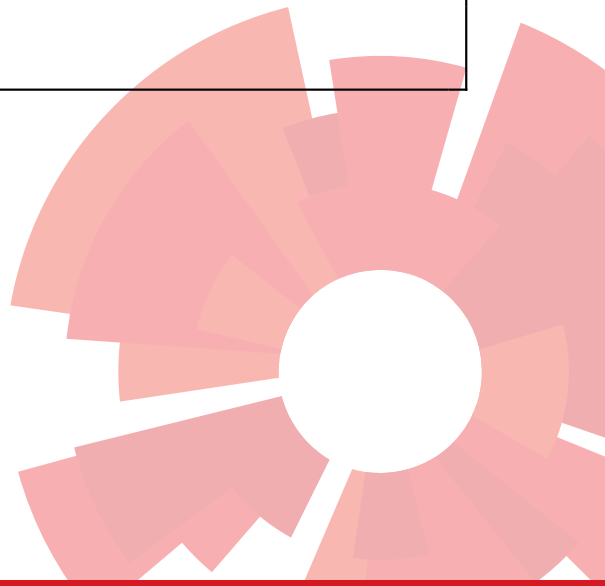
JSON

```
>>> json.loads(js)
{'boolean': True,
 'key1': 3,
 'key2': 43.2,
 'list': [1, 2, 3],
 'test': None}
```



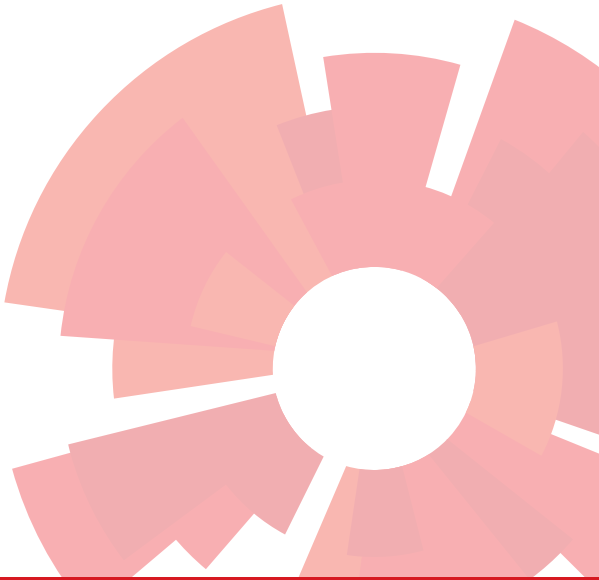
JSON

```
>>> json.dumps('''{
...     'test': 3,
...     5: True,
...     4.3: None
... }''')
('"{\\n    \'test\': 3,\\n    5: True,\\n    '
 '4.3: None\\n}"')
```



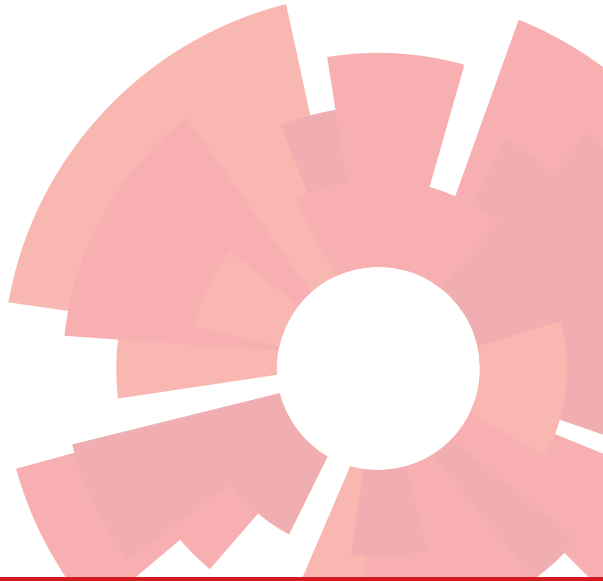
Serialization

- Translating data structures into a format so that they can be recreated again
- JSON can be used to serialize a subset of Python's data structures
  - Plain text serialization



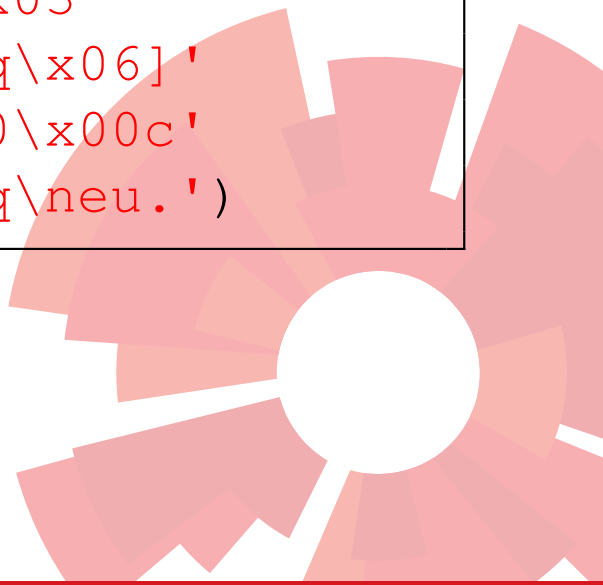
Pickle

- Python specific serialization
  - Binary serialization
- Can serialize
  - Most/all Python data structures
  - Functions
  - Classes



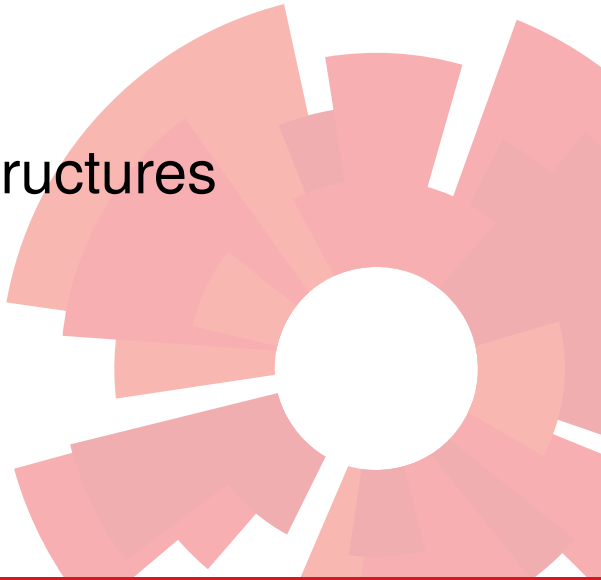
Pickle

```
>>> import pickle
>>> pickle.dumps({
...     'a': {1,2,3},
...     'b': [1,2,3],
...     'c': [True, None, (1,2,3)]
... })
(b'\x80\x03}q\x00(X\x01\x00\x00\x00aq\x01cbuilt'
 b'ins\nset\nq\x02]q\x03(K\x01K\x02K\x03'
 b'e\x85q\x04Rq\x05X\x01\x00\x00\x00bq\x06]'
 b'q\x07(K\x01K\x02K\x03eX\x01\x00\x00\x00c'
 b'q\x08]q\t(\x88NK\x01K\x02K\x03\x87q\neu.')
```



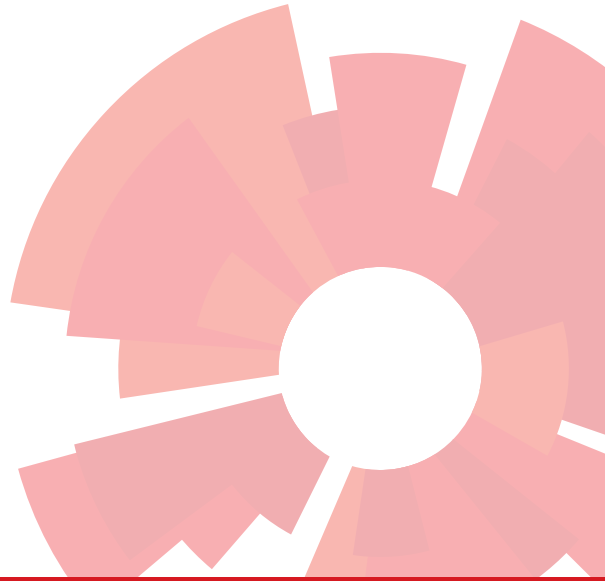
Pickle vs. JSON

- JSON
  - Plain text
  - Can only serialize a small part of Python's data structures
  - Widely used
- Python
  - Binary
  - Can serialize most of Python's data structures
  - Not (or rarely) used outside Python



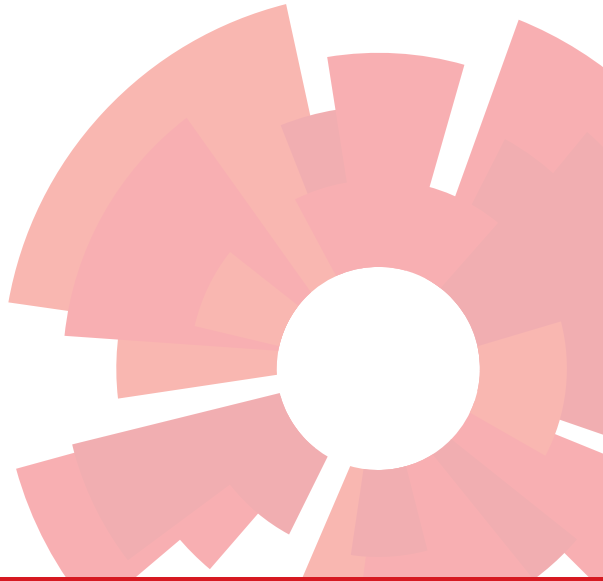
XML

- Plain text
- `xml.etree.ElementTree`



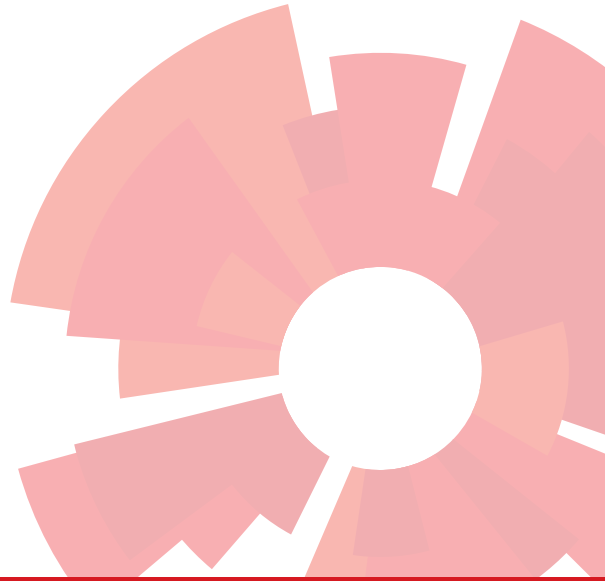
sqlite3

- SQL database in a single file



CSV

- Comma Separated Value
- Tabular data



HTTP requests

- Fetch data over HTTP (and HTTPS and FTP)