Coming Up Next…

- Getting started with the Fake News Challenge data

Coming Up Next…

- Getting started with the Fake News Challenge data
- Introduction to the Fake News Challenge

Coming Up Next…

- Getting started with the Fake News Challenge data
- Introduction to the Fake News Challenge
- Rule-based models (our first algorithm!)

Coming Up Next…

- Getting started with the Fake News Challenge data
- Introduction to the Fake News Challenge
- Rule-based models (our first algorithm!)
- Discussion of evaluation metrics

Coming Up Next…

- Getting started with the Fake News Challenge data
- Introduction to the Fake News Challenge
- Rule-based models (our first algorithm!)
- Discussion of evaluation metrics
- Building a more sophisticated classifier

Coming Up Next…

- **Getting started with the Fake News Challenge data**
- Introduction to the Fake News Challenge
- Rule-based models (our first algorithm!)
- Discussion of evaluation metrics
- Building a more sophisticated classifier

Loading the Data

AI4ALL Princeton: NLP Group

Review of Python and Pandas

pandas.DataFrame.head

`DataFrame.`**head**(*n=5*) [source]

Return the first *n* rows.

This function returns the first *n* rows for the object based on position. It is useful for quickly testing if your object has the right type of data in it.

Parameters:	**n** *: int, default 5* Number of rows to select.
Returns:	**obj_head** *: type of caller* The first *n* rows of the caller object.

See also:

pandas.DataFrame.tail
Returns the last *n* rows.

Examples

```
>>> df = pd.DataFrame({'animal':['alligator', 'bee', 'falcon', 'lion',
...                    'monkey', 'parrot', 'shark', 'whale', 'zebra']})
>>> df
      animal
0  alligator
1        bee
2     falcon
3       lion
4     monkey
5     parrot
6      shark
7      whale
8      zebra
```

Some helpful commands:

- pd.read_csv()
- pd.unique(series)

- df.columns
- df.shape
- df.head()
- df[column_name]

- len(array)

Some helpful commands:

- pd.read_csv()
- pd.unique(series)

- df.columns
- df.shape
- df.head()
- df[column_name]

- len(array)

With a partner,
look up these functions
and discuss what they do

Now let’s get to the FNC data!

1. Go to https://github.com/cchen23/AI4ALL_NLP_student
2. Download 2018_07_24.zip
3. Unzip the file
4. Open Terminal
5. `source ~/miniconda3/bin/activate`
6. `jupyter notebook`
7. Open Day2_NLP_LoadingTheData.ipynb

Feedback Form

http://bit.ly/2uFuaPl