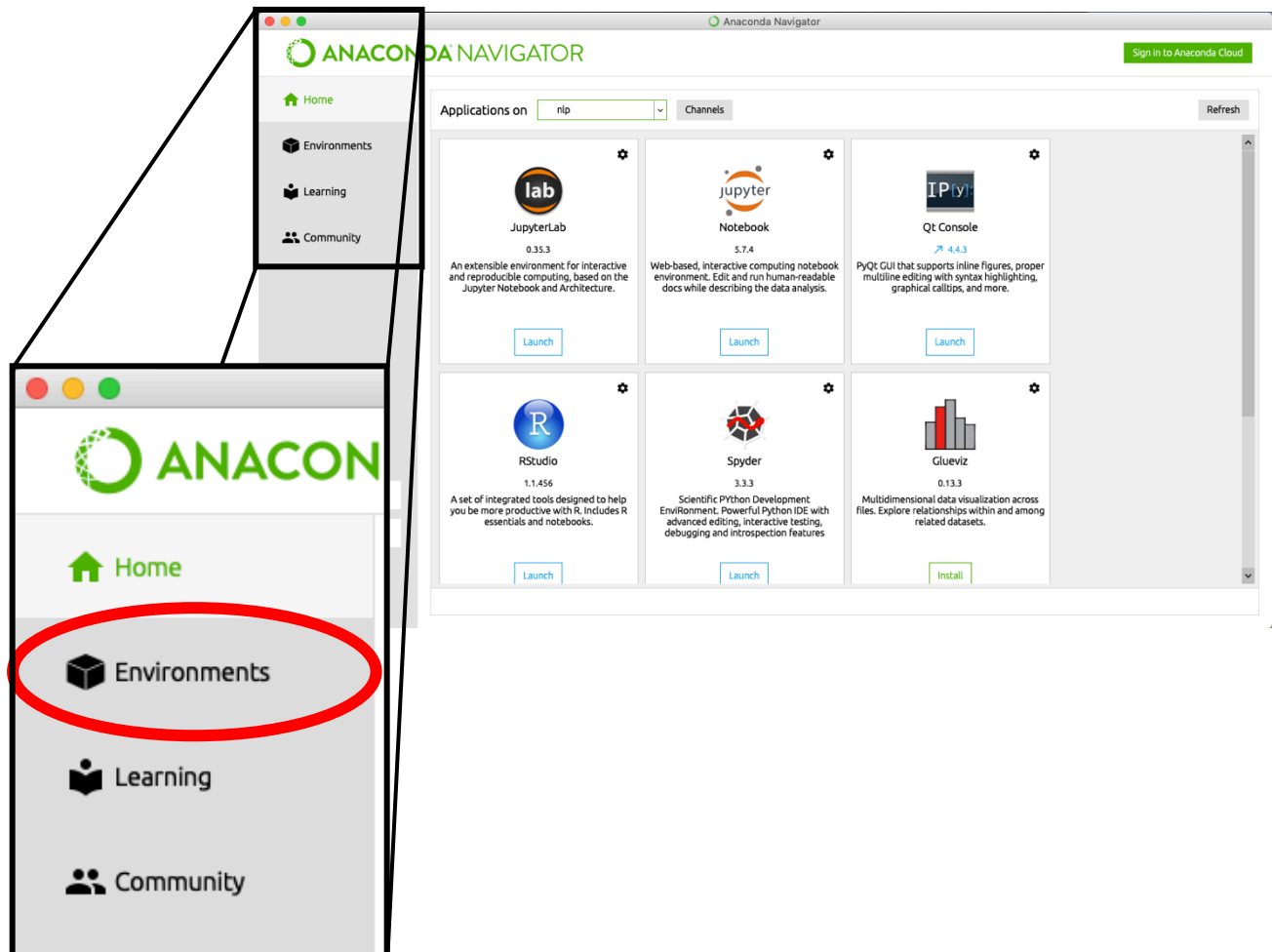


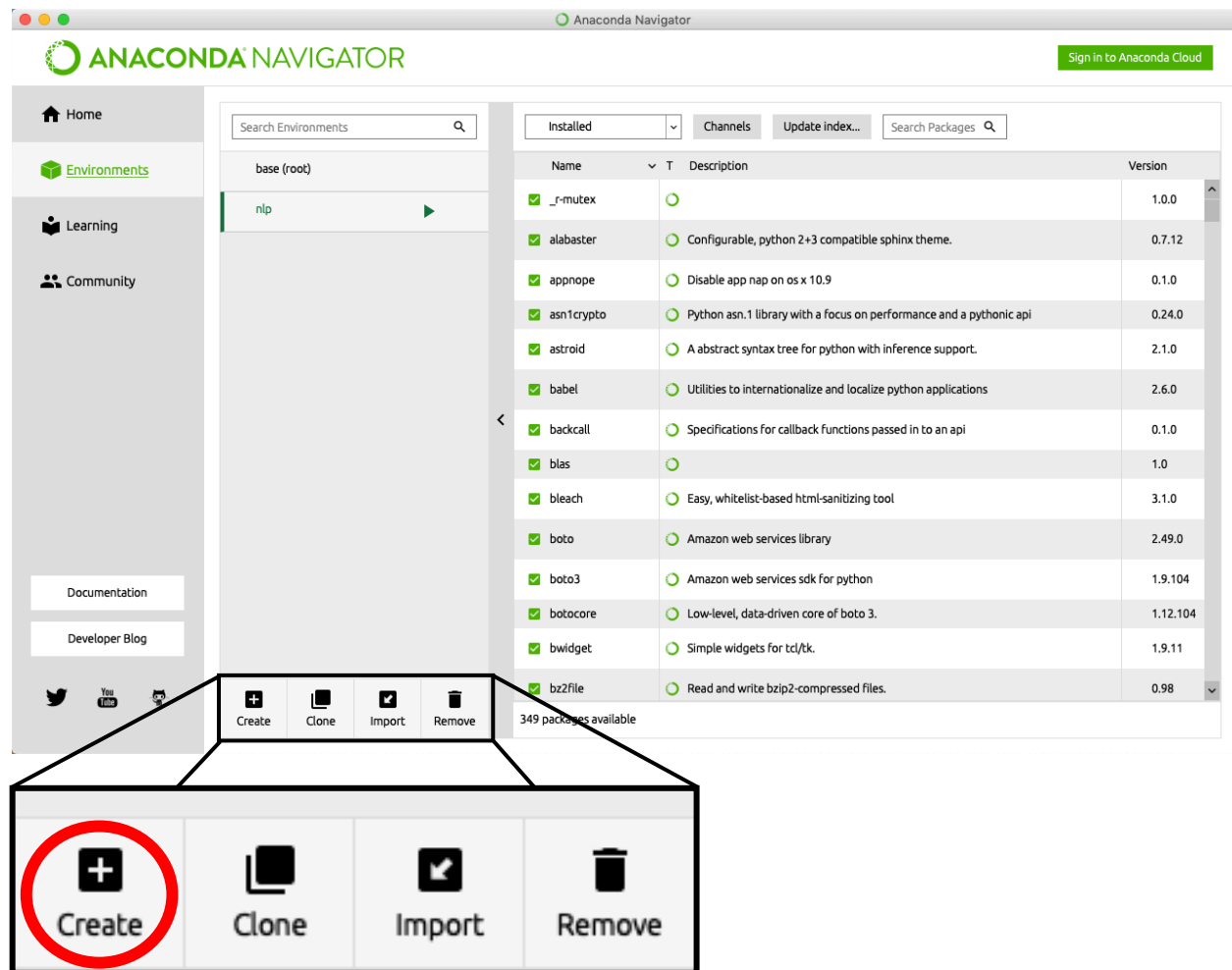
Hayley's Recommendations for NLP

Hi guys! I'm super excited to work on this project with you. I realized as I was looking through all of my files that I had an incredibly frustrating time getting everything working and installed when I was starting this. So, I'm writing this (hopefully) helpful document to get an environment set up. Let's get started!

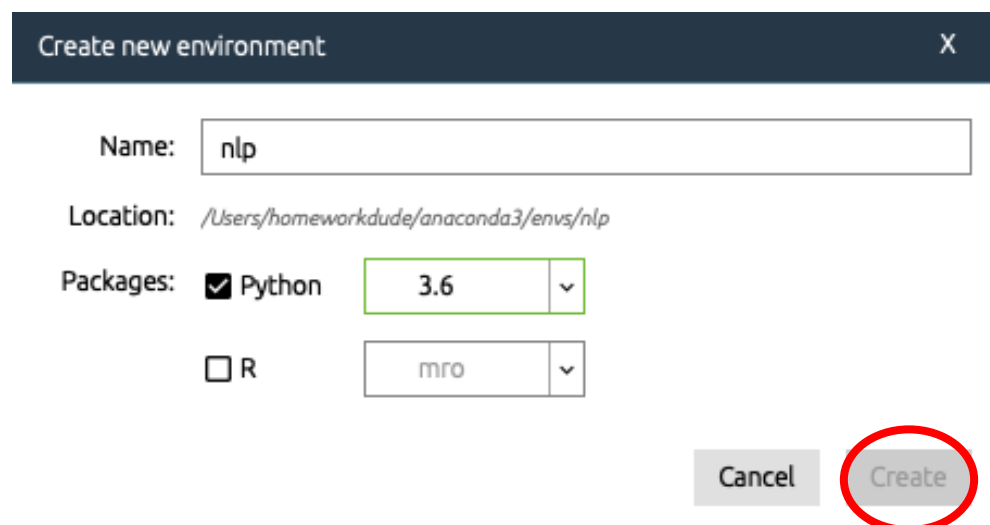
First things first, let's open Anaconda. Mine looks like this! We're going to go to the Environments tab on the left hand side.



Here we are! We are going to make a new environment called 'nlp'.



Down at the bottom of your screen you will see a create button. Select that. A window like the one shown below will pop up. Set the python version to be 3.6 and select create.



After creating your environment, open up the Anaconda Prompt. Personally, I don't have the Anaconda Prompt, but I modified my terminal to do the same thing. Here we are going to install NLTK, TextBlob, spaCy, and genism. Enter the command `conda activate nlp`, shown below. Notice how the name to the left of the drive name changes from base to nlp.

```
[(base) Hayleys-MacBook-Pro:~ homeworkdude$ conda activate nlp  
(nlp) Hayleys-MacBook-Pro:~ homeworkdude$
```

Ok awesome, we're inside nlp. Now we're going to start off by installing NLTK. All commands done in this space will start with conda. So to install NLTK, we are going to give the command `conda install nltk`.

```
[(nlp) Hayleys-MacBook-Pro:~ homeworkdude$ conda install nltk  
Collecting package metadata: done  
Solving environment: done  
  
## Package Plan ##  
  
environment location: /Users/homeworkdude/anaconda3/envs/nlp  
  
added / updated specs:  
- nltk  
  
The following packages will be downloaded:  
  
package | build | size  
-----|-----|-----  
nltk-3.4 | py36_1 | 2.1 MB  
-----|-----|-----  
Total: | | 2.1 MB  
  
The following NEW packages will be INSTALLED:  
  
nltk pkgs/main/osx-64::nltk-3.4-py36_1  
  
Proceed ([y]/n)? y  
  
Downloading and Extracting Packages  
nltk-3.4 | 2.1 MB | ##### | 100%  
Preparing transaction: done  
Verifying transaction: done  
Executing transaction: done  
(nlp) Hayleys-MacBook-Pro:~ homeworkdude$
```

It will verify that installing the package is ok, to which you just type y and hit enter and you're on your way! Now that NLTK is installed, we can get to the meatier packages. Next up, TextBlob.

To install TextBlob, we need two commands. First, double check to verify that pip is installed in your environment by running `pip --version`.

```
((nlp) Hayleys-MacBook-Pro:~ homeworkdude$ pip --version
pip 19.0.3 from /Users/homeworkdude/anaconda3/envs/nlp/lib/python3.6/site-packages/pip (python 3.6)
```

Awesome, pip is installed and up to date. If yours is not, run `conda install pip`. Now that pip is in the environment, we can run the install command for TextBlob, `pip install -U textblob`.

```
((nlp) Hayleys-MacBook-Pro:~ homeworkdude$ pip install -U textblob
Requirement already up-to-date: textblob in ./anaconda3/envs/nlp/lib/python3.6/site-packages (0.15.3)
Requirement already satisfied, skipping upgrade: nltk>=3.1 in ./anaconda3/envs/nlp/lib/python3.6/site-packages (from textblob) (3.4)
Requirement already satisfied, skipping upgrade: six in ./anaconda3/envs/nlp/lib/python3.6/site-packages (from nltk>=3.1->textblob) (1.12.0)
Requirement already satisfied, skipping upgrade: singledispatch in ./anaconda3/envs/nlp/lib/python3.6/site-packages (from nltk>=3.1->textblob) (3.4.0.3)
((nlp) Hayleys-MacBook-Pro:~ homeworkdude$
```

My print out will look different from y'all's on all of these because I have them all installed already. Anyway, TextBlob is now installed. The second command will download the corpus that we will be using. Run the command `python -m textblob.download_corpora`.

```
((nlp) Hayleys-MacBook-Pro:~ homeworkdude$ python -m textblob.download_corpora
[nltk_data] Downloading package brown to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package brown is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-date!
[nltk_data] Downloading package conll2000 to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package conll2000 is already up-to-date!
[nltk_data] Downloading package movie_reviews to
[nltk_data]   /Users/homeworkdude/nltk_data...
[nltk_data]   Package movie_reviews is already up-to-date!
Finished.
((nlp) Hayleys-MacBook-Pro:~ homeworkdude$
```

Yay TextBlob!! Now on to spaCy. The first command is `conda install -c conda-forge spacy`.

```
[(nlp) Hayleys-MacBook-Pro:~ homeworkdude$ conda install -c conda-forge spacy
Collecting package metadata: done
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: /Users/homeworkdude/anaconda3/envs/nlp
```

```
added / updated specs:
- spacy
```

```
The following packages will be downloaded:
```

package	build		
cymem-2.0.2	py36h04f5b5a_0	30 KB	
cytoolz-0.9.0.1	py36h1de35cc_1001	358 KB	conda-forge
dill-0.2.9	py36_0	113 KB	conda-forge
msgpack-numpy-0.4.3.2	py_0	8 KB	conda-forge
msgpack-python-0.6.1	py36h04f5b5a_0	82 KB	conda-forge
murmurhash-1.0.0	py36h0a44026_0	15 KB	conda-forge
openssl-1.1.1b	h1de35cc_1	3.5 MB	conda-forge
plac-0.9.6	py_1	18 KB	conda-forge
preshed-2.0.1	py36h0a44026_0	63 KB	
regex-2018.01.10	py36h1de35cc_1000	333 KB	conda-forge
spacy-2.0.18	py36h0a44026_1000	51.4 MB	conda-forge
thinc-6.12.1	py36h1702cab_1000	1.3 MB	conda-forge
toolz-0.9.0	py_1	42 KB	conda-forge
tqdm-4.31.1	py_0	40 KB	conda-forge
ujson-1.35	py36h1de35cc_1001	24 KB	conda-forge
wrapt-1.10.11	py36h1de35cc_1001	41 KB	conda-forge
Total:		57.3 MB	

```
The following NEW packages will be INSTALLED:
```

cymem	pkgs/main/osx-64::cymem-2.0.2-py36h04f5b5a_0
cytoolz	conda-forge/osx-64::cytoolz-0.9.0.1-py36h1de35cc_1001
dill	conda-forge/osx-64::dill-0.2.9-py36_0
msgpack-numpy	conda-forge/noarch::msgpack-numpy-0.4.3.2-py_0
msgpack-python	conda-forge/osx-64::msgpack-python-0.6.1-py36h04f5b5a_0
murmurhash	conda-forge/osx-64::murmurhash-1.0.0-py36h0a44026_0
plac	conda-forge/noarch::plac-0.9.6-py_1
preshed	pkgs/main/osx-64::preshed-2.0.1-py36h0a44026_0
regex	conda-forge/osx-64::regex-2018.01.10-py36h1de35cc_1000
spacy	conda-forge/osx-64::spacy-2.0.18-py36h0a44026_1000
thinc	conda-forge/osx-64::thinc-6.12.1-py36h1702cab_1000
toolz	conda-forge/noarch::toolz-0.9.0-py_1
tqdm	conda-forge/noarch::tqdm-4.31.1-py_0
ujson	conda-forge/osx-64::ujson-1.35-py36h1de35cc_1001

```
The following packages will be UPDATED:
```

certifi	pkgs/main::certifi-2018.11.29-py36_0 --> conda-forge::certifi-2018.11.29-py36_1000
openssl	pkgs/main::openssl-1.1.1b-h1de35cc_0 --> conda-forge::openssl-1.1.1b-h1de35cc_1

```
The following packages will be SUPERSEDED by a higher-priority channel:
```

ca-certificates	pkgs/main::ca-certificates-2019.1.23-0 --> conda-forge::ca-certificates-2018.11.29-ha4d7672_0
wrapt	pkgs/main::wrapt-1.11.1-py36h1de35cc_0 --> conda-forge::wrapt-1.10.11-py36h1de35cc_1001

```
Proceed ([y]/n)? y
```

```

Downloading and Extracting Packages
tqdm-4.31.1      | 40 KB      | ##### | 100%
wrapt-1.10.11    | 41 KB      | ##### | 100%
msgpack-numpy-0.4.3 | 8 KB      | ##### | 100%
toolz-0.9.0      | 42 KB      | ##### | 100%
thinc-6.12.1     | 1.3 MB     | ##### | 100%
spacy-2.0.18     | 51.4 MB    | ##### | 100%
msgpack-python-0.6.1 | 82 KB     | ##### | 100%
murmurhash-1.0.0  | 15 KB      | ##### | 100%
cytoolz-0.9.0.1  | 358 KB     | ##### | 100%
preshed-2.0.1    | 63 KB      | ##### | 100%
cymem-2.0.2      | 30 KB      | ##### | 100%
ujson-1.35       | 24 KB      | ##### | 100%
regex-2018.01.10 | 333 KB     | ##### | 100%
dill-0.2.9       | 113 KB     | ##### | 100%
plac-0.9.6       | 18 KB      | ##### | 100%
openssl-1.1.1b   | 3.5 MB     | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
(nlp) Hayleys-MacBook-Pro:~ homeworkdude$

```

Next we are going to download the corpora for spaCy. The command for this one is `python -m spacy download en`.

```

(nlp) Hayleys-MacBook-Pro:~ homeworkdude$ python -m spacy download en
Requirement already satisfied: en_core_web_sm==2.0.0 from https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.0.0/en_core_web_sm-2.0.0.tar.gz#egg=en_core_web_sm==2.0.0 in
./anaconda3/envs/nlp/lib/python3.6/site-packages (2.0.0)

Linking successful
/Users/homeworkdude/anaconda3/envs/nlp/lib/python3.6/site-packages/en_core_web_sm
-->
/Users/homeworkdude/anaconda3/envs/nlp/lib/python3.6/site-packages/spacy/data/en

You can now load the model via spacy.load('en')

(nlp) Hayleys-MacBook-Pro:~ homeworkdude$

```

Onto the last one, gensim. I had the most trouble getting spacy to work on my machine, so we should be past the absolute worst of this process. The command for installing gensim is `conda install -c conda-forge gensim`. Boom done. Super easy.

```

(nlp) Hayleys-MacBook-Pro:~ homeworkdude$ conda install -c conda-forge gensim
Collecting package metadata: done
Solving environment: done

# All requested packages already installed.

(nlp) Hayleys-MacBook-Pro:~ homeworkdude$

```

One last package that I accidentally forgot to mention. This one is fun, it creates interactive bubble charts for topic modeling. This one is called pyLDAvis. The install command is `conda install -c conda-forge pyldavis`.

```
[(nlp) Hayleys-MacBook-Pro:~ homeworkdude$ conda install -c conda-forge pyldavis  
Collecting package metadata: done  
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: /Users/homeworkdude/anaconda3/envs/nlp
```

```
added / updated specs:  
- pyldavis
```

```
The following packages will be downloaded:
```

package	build		
funcy-1.11	py_0	26 KB	conda-forge
future-0.17.1	py36_1000	699 KB	conda-forge
joblib-0.13.2	py_0	180 KB	conda-forge
numexpr-2.6.9	py36h1702cab_1000	128 KB	conda-forge
pandas-0.24.1	py36h0a44026_0	10.1 MB	conda-forge
pyldavis-2.1.2	py_0	81 KB	conda-forge
Total:		11.2 MB	

```
The following NEW packages will be INSTALLED:
```

funcy	conda-forge/noarch::funcy-1.11-py_0
future	conda-forge/osx-64::future-0.17.1-py36_1000
joblib	conda-forge/noarch::joblib-0.13.2-py_0
numexpr	conda-forge/osx-64::numexpr-2.6.9-py36h1702cab_1000
pandas	conda-forge/osx-64::pandas-0.24.1-py36h0a44026_0
pyldavis	conda-forge/noarch::pyldavis-2.1.2-py_0

```
Proceed ([y]/n)? y
```

```
Downloading and Extracting Packages
```

joblib-0.13.2	180 KB	#####	100%
future-0.17.1	699 KB	#####	100%
numexpr-2.6.9	128 KB	#####	100%
funcy-1.11	26 KB	#####	100%
pandas-0.24.1	10.1 MB	#####	100%
pyldavis-2.1.2	81 KB	#####	100%

```
Preparing transaction: done
```

```
Verifying transaction: done
```

```
Executing transaction: done
```

```
(nlp) Hayleys-MacBook-Pro:~ homeworkdude$
```

We are now officially ready to perform natural language processing in python! Yay! Eventually we will probably have to add scikit learn in later if that's what we use for machine learning, *but* we will cross that bridge when we get to it. I hope this little tutorial helped, if not, well I just wasted 2 hours 🙄 oh well. C'est la vie, on to the next project! 🐼

Hayley

Links to documentation:

<https://anaconda.org/conda-forge/pyldavis>

<https://radimrehurek.com/gensim/install.html>

<https://spacy.io/usage/>

<https://textblob.readthedocs.io/en/dev/install.html>

<https://www.nltk.org/install.html>