Worksheet #1: After running the cp command, what file(s) are now in your home directory? There should be at least two: a ".sh" file, and a ".py" file.
$ ls
ngram-job.py spark-run.sh

Worksheet #2: What is the name of the last file in the listing for HFS folder /var/si618f17?
-rw-r----- 3 deahanyu si618f17 2292130515 2017-09-18 20:46 /var/si618f17/
yelp_academic_dataset_review.json

Worksheet #3: What year was Einstein first mentioned (as a noun) in Google Books data?
einstein_NOUN 1921 4 4
- Year 1921 was first mentioned.

Worksheet #4: After the Spark job completes, what are three files listed in your Hadoop File System output directory ./ngrams-out?
Found 3 items
-rw-r----- 3 lizeyu hadoop 0 2017-09-19 15:50 ngrams-out/_SUCCESS
-rw-r----- 3 lizeyu hadoop 5540 2017-09-19 15:50 ngrams-out/part-00000
-rw-r----- 3 lizeyu hadoop 5492 2017-09-19 15:50 ngrams-out/part-00001

Worksheet #5: What were the average word lengths observed in books from the years 1520, 1597, and 1598 ?
(Year, Avg word length):
(1520, 10.84)
(1597, 8.6)
(1598, 9.832)

6. Bonus Challenge