Homework 3: the Terminal & Beginning Python

LIN 301: Computation for Linguists

Dr. Andrew M. Byrd

Part 1: Terminal (Setup)

1. Open your Terminal in RStudio.

2. Navigate (`cd`) into your course repo folder (e.g., `LIN_301`).

3. Make a new folder called `hw3_files`:

4. Download the file **shakespeare.txt** from Project Gutenberg into that folder (use `curl`). Tje file is located at this address: https://www.gutenberg.org/cache/epub/100/pg100.txt.

5. Confirm the file is in your folder with the `ls` command.

Part 2: Python (Lists and Dictionaries)

Create a new Quarto doc called **hw3.qmd** with a Python code chunk.

Load the File

```
with open("hw3_files/shakespeare.txt", "r", encoding="utf-8") as f:
    text = f.read()
```

Make a List

- Using the method `.split()[:1000]`, take the first 1000 words (in the order they appear) and store them in a list called `first_words`.

- Print the list.
- Print the 500th item in the list.

Convert to a Set

- Another type of container (like lists & dictionaries) in Python is a **set**, which is like a dictionary with just keys (and no values). It's unordered and allows no duplicates. This will help us eliminate any duplicates.

```
unique_words = list(set(first_words))
print(unique_words)
```

Narrowing down the List

- Now pick **10 of the words** from your `first_words` list. You should pick super common ones (like 'the') and rarer ones, too.
- Manually place these 10 words in a new list called `chosen_ones`.

Make a Dictionary

- use the method '.split()' to separate *all* words into a new list.
- Use `.count()` to see how many times each appears in the text.

- Create a dictionary called `word_counts` that pairs each chosen word with its number.

Example:

```
word_counts = {
    "The": text.count("The"),
    "Project": text.count("Project"),
    "Shakespeare": text.count("Shakespeare")
}
print(word_counts)
```

Part 3: Reflection

At the bottom of your **hw3.qmd**, write a short (3–5 sentence) reflection:

- What was easiest about this assignment?

- What was hardest?

- How might a linguist use lists and dictionaries for language data?

Submission

- Upload **hw3.qmd** and your **hw3_files/shakespeare.txt** file to Canvas.