# Exercises

Find the number of words (in `/usr/share/dict/words`) that contain at least three `a`s and don't have a `'s` ending. What are the three most common last two letters of those words? `sed`'s `y` command, or the `tr` program, may help you with case insensitivity. How many of those two-letter combinations are there? And for a challenge: which combinations do not occur?

*There are 846 words that contain at least three `a`s and not ending in `'s`.*

```
~$ cat /usr/share/dict/words \
> | egrep -i '(a.*?){3,}' \
> | grep -vi "'s$" \
> | wc -l
846
```

`an, ns, as` *are the most common last two letters*

```
~$ cat /usr/share/dict/words \
> | egrep -i '(a.*?){3,}' \
> | grep -vi "'s$" \
> | sed -E 's/.*(..)/\1/' \
> | sort | uniq -c | sort -nr | head -3 \
> | awk '{print $2}' | paste -sd,
an,ns,as
```

```
~$ cat /usr/share/dict/words \
> | egrep -i '(a.*?){3,}' \
> | grep -vi "'s$" \
> | sed -E 's/.*(..)/\1/' \
> | sort | uniq -c | sort -nr | head -3
    101 an
     63 ns
     54 as
```

*A total of 218 words end in either* `an`, `ns`, *or* `as`.

```
~$ cat /usr/share/dict/words \
> | egrep -i '(a.*?){3,}' \
> | grep -vi "'s$" \
> | sed -E 's/.*(..)/\1/' \
> | sort | uniq -c | sort -nr | head -3 \
> | awk '{print $1}' | paste -sd+ | bc
218
```

To do in-place substitution it is quite tempting to do something like `sed s/REGEX/SUBSTITUTION/ input.txt > input.txt`. However this is a bad idea, why? Is this particular to `sed`? Use `man sed` to find out how to accomplish this.

Simply using `sed s/REGEX/SUBSTITUTION/ input.txt > input.txt` to do in-place substitution risks corrupting the input file when the operation gets disrupted or when something goes wrong with the execution. This isn't particular to `sed`. We can use the `-i` option in `sed` to edit files in-place and save backups of the original file by suffixing it with the extension.

Find your average, median, and max system boot time over the last ten boots. Use `journalctl` on Linux and `log show` on macOS, and look for log timestamps near the beginning and end of each boot.

```
# Finding patterns in startup logs
~$ journalctl | grep 'Startup finished in' | head
Nov 01 09:45:01 ramzel-Inspiron-14-3467 systemd[1]: Startup finished
in 4.004s (kernel) + 4min 9.897s (userspace) = 4min 13.901s.
Nov 02 08:34:04 ramzel-Inspiron-14-3467 systemd[1238]: Startup
finished in 773ms.
Nov 02 08:34:28 ramzel-Inspiron-14-3467 systemd[1]: Startup finished
in 4.005s (kernel) + 1min 16.462s (userspace) = 1min 20.468s.
...
# Using keyword 'userspace' to pull up boot time logs
~$ journalctl | grep userspace | tail
Apr 12 08:38:26 ramzel-Inspiron-14-3467 systemd[1]: Startup finished
in 6.415s (firmware) + 4.995s (loader) + 4.206s (kernel) + 1min
```

```
21.085s (userspace) = 1min 36.704s.
Apr 17 09:23:30 ramzel-Inspiron-14-3467 systemd[1]: Startup finished
in 6.490s (firmware) + 4.941s (loader) + 6.013s (kernel) + 5min
25.813s (userspace) = 5min 43.259s.
~$ journalctl | grep userspace | tail \
# Getting only the minutes and seconds from the line
> | sed -E 's/^.*= (.*)s/\1/' | sed 's/min//' \
1 36.704.
5 43.259.
2 35.679.
1 35.351.
# Converting all into seconds
> | awk {'print $1 * 60 + $2'} \
96.704
343.259
155.679
...
# Print boot time summary statistics (in seconds)
> | R --slave -e 'x <- scan(file="stdin", quiet=TRUE); summary(x)'
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  88.25   93.33   96.03  192.09  188.03  664.84
```

Look for boot messages that are *not* shared between your past three reboots (see `journalctl`'s `-b` flag). Break this task down into multiple steps. First, find a way to get just the logs from the past three boots. There may be an applicable flag on the tool you use to extract the boot logs, or you can use `sed '0,/STRING/d'` to remove all lines previous to one that matches `STRING`. Next, remove any parts of the line that *always* varies (like the timestamp). Then, de-duplicate the input lines and keep a count of each one (`uniq` is your friend). And finally, eliminate any line whose count is 3 (since it *was* shared among all the boots).

```
~$ man journalctl
# Creating empty file for boot logs
~$ touch boot_logs
# Write to file the logs for the past three boots
# Remove duplicate lines and redundant data
~$ journalctl -b | sed -E 's/^.*ramzel-Inspiron-14-3467 (.*)$/\1/' |
sort -u >> boot_logs
```

```
~$ journalctl -b -1 | sed -E 's/^.*ramzel-Inspiron-14-3467 (.*)$/\1/'
| sort -u >> boot_logs
~$ journalctl -b -2 | sed -E 's/^.*ramzel-Inspiron-14-3467 (.*)$/\1/'
| sort -u >> boot_logs
# Leave out only the gist of the lines as much as possible
# Removing data that always varies (like timestamps)
~$ cat boot_logs | sed -E 's/^.*: //' \
> | sed -E 's/[^A-Za-z]//g' \
> | sort | uniq -c \
# Remove shared lines between past three boots
> | awk '{print $1}' | grep 3 | wc -l
1815
> | awk '$1 < 3 { print $2 }' > uniq_messages
$ shuf -n5 uniq_messages
TTYptsPWDhomeramzelUSERrootCOMMANDusrbinaptinstallrubyfull
FailedparsingCertificate
AprSERVICEWORKERfetchingassethttpsaslackedgecombvclientlazycomponents
aedcebminjs
StoppingUpdateUTMPaboutSystemBootShutdown
```