Group Assignment 1

Group

2022-10-21

```
## Loading required package: NLP
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

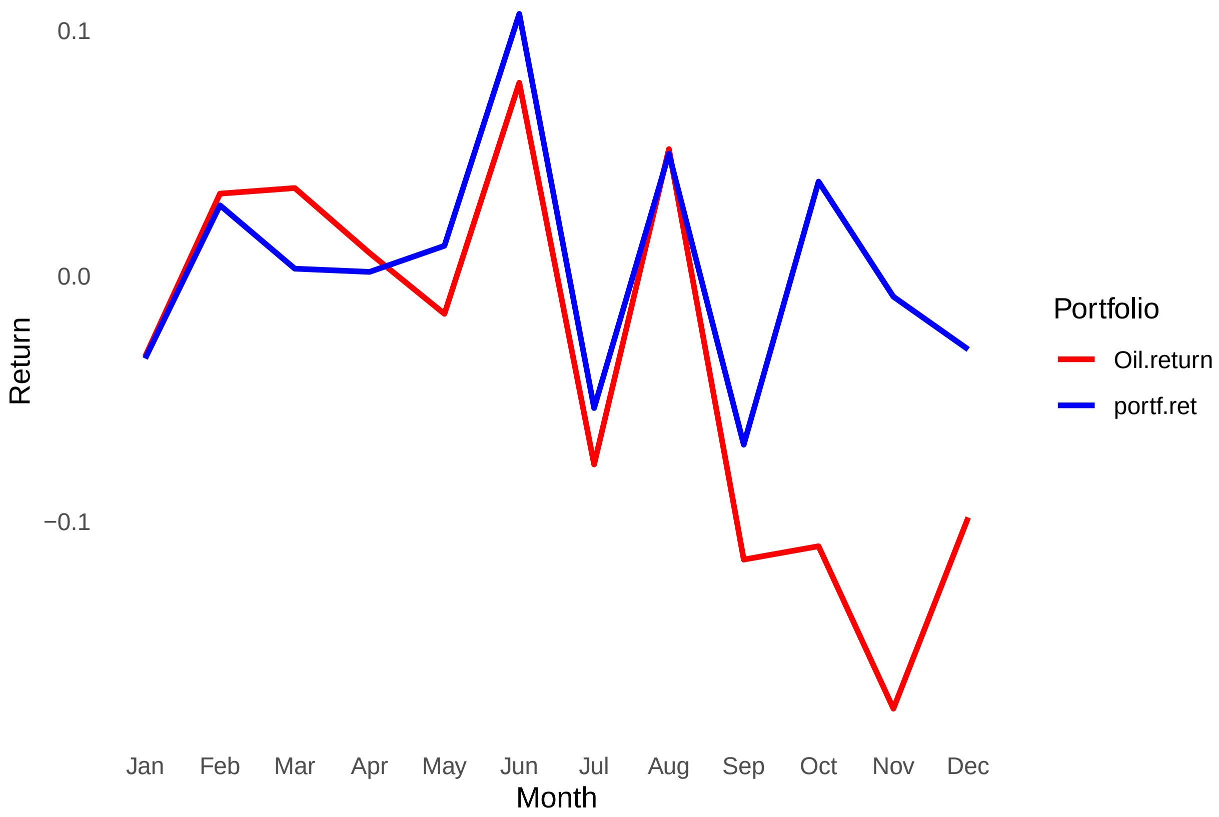
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:NLP':
##
##     annotate
```

```
## Selecting by ll
```

```
##                             ll
## proved_reserves      1004.0300
## working_interest      991.5506
## hydraulic_fracturing  566.1345
## proved_undeveloped    516.7899
## gross_acres           462.5419
## estimated_proved      444.2142
## shale_play            387.4694
## undeveloped_reserves  366.5558
## natural_production    319.3342
## reserves_december     285.7961
```



Evaluate the performance of your algorithm. Compute the Root Mean Squared Error (just google the definition) between both portfolios.

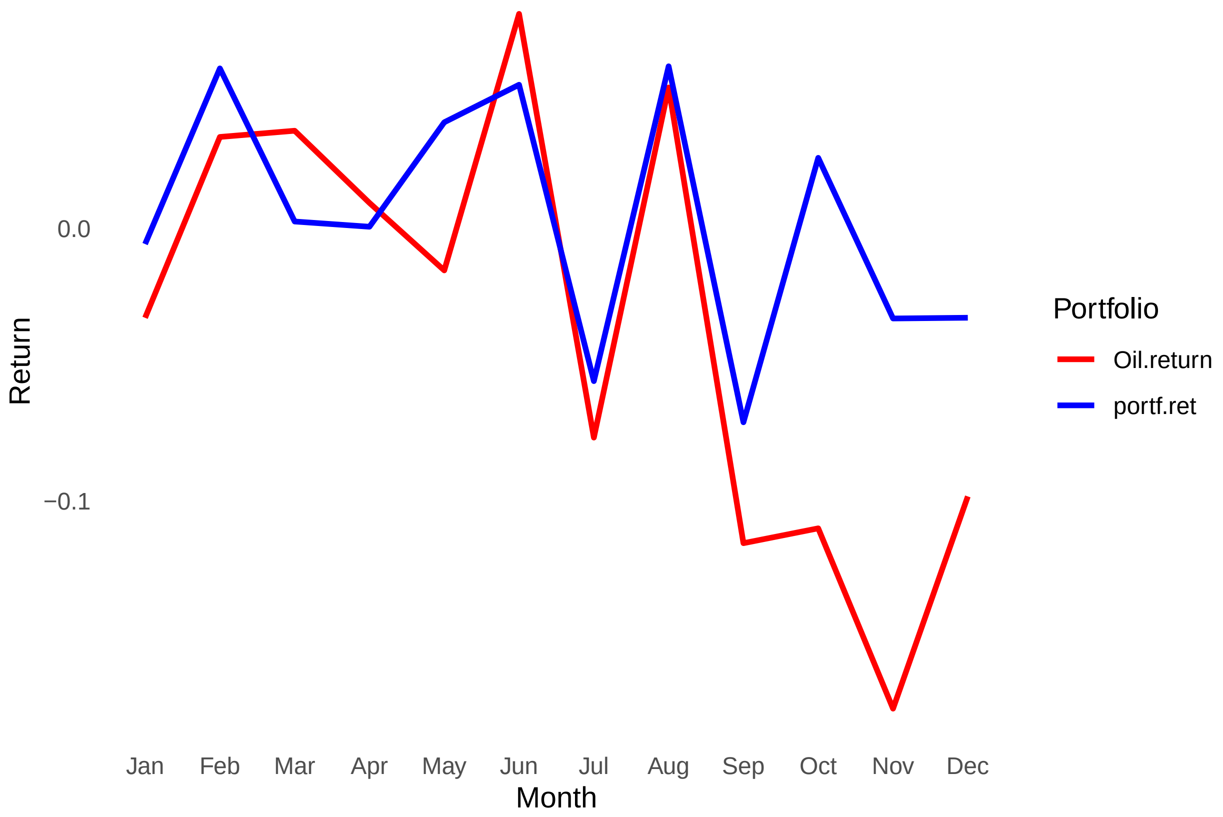
```
##        RMSE
## 0.005029665
```

Evaluate whether using uni-grams (versus bi-grams) performs better/worse? For the remainder of steps, use exactly the same approach. An easy way to implement this analysis is by recycling your current code skipping the step in which you transform the corpus into bigrams.

Uni-grams

```
## Selecting by ll
```

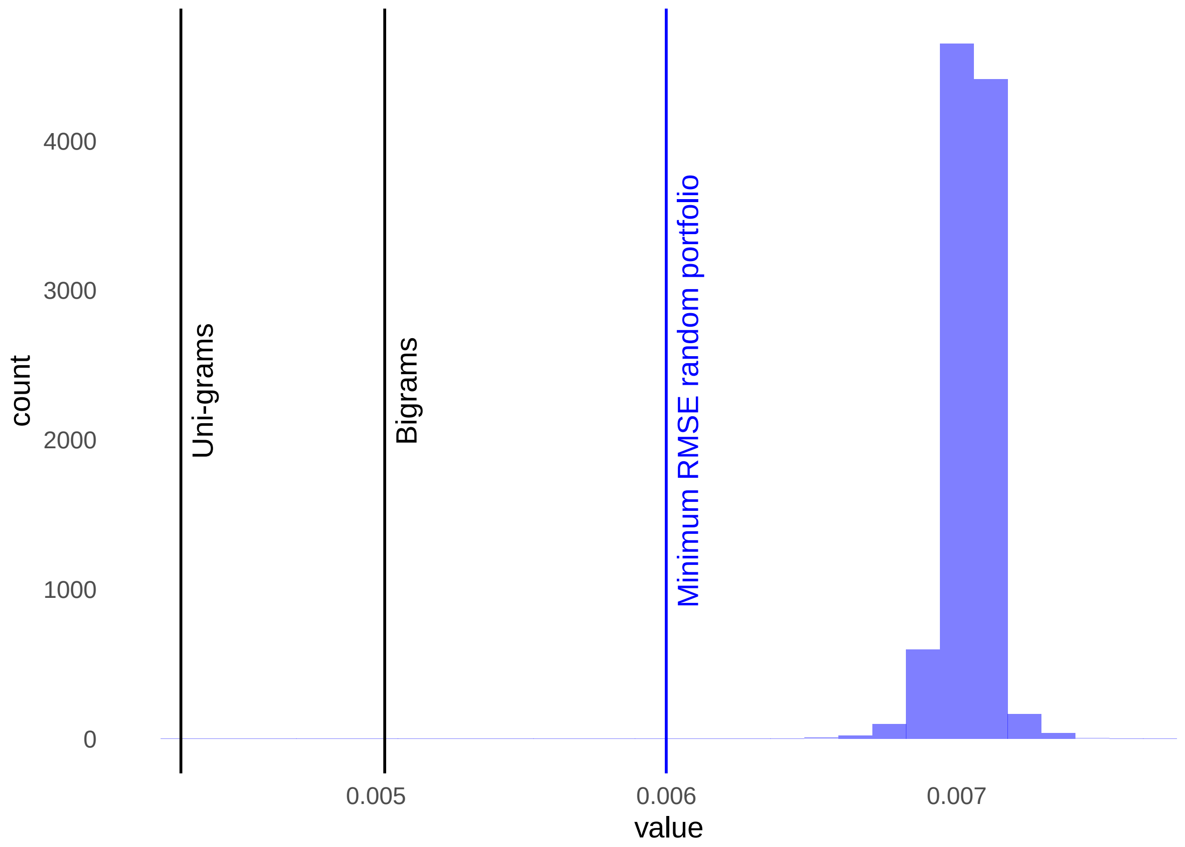
```
##                   ll
## natural     2906.320
## wells       2311.355
## production  2059.049
## proved      1989.841
## drilling    1861.884
## reserves    1633.491
## block       1458.118
## field       1440.347
## exploration 1386.708
## acres       1007.866
```



```
##        RMSE
## 0.004327684
```

- [Optional] Does this approach of constructing tracking portfolios work? Knowing the value of the RMSE is not necessarily enough to assess performance against alternative portfolio selections. Construct 10.000 random portfolios of similar number of firms and compute the corresponding RMSE. Display it as a histogram. Does the text approach do a better/worse job?

```
## Warning in rm(replicating.portf.rand): object 'replicating.portf.rand' not found
```


count
4000
3000
2000
1000
0
Uni-grams
Bigrams
Minimum RMSE random portfolio
0.005
0.006
0.007
value