

# Trabalho 1

Aluno: Cristiane de Paula Oliveira - Matrícula: 261424  
IF-UFRGS

29 de março de 2016

## Resumo

Neste trabalho está relatado as estatísticas para cinco capítulos do romance Moby Dick de Herman Melville: *Chapter 1 - Loomings*[1]; *Chapter 41 - Moby Dick*[2]; *Chapter 42 - The Whiteness of The Whale*[3]; *Chapter 132 - The Symphony*[4] e *Epilogue*[5].

*“Call me Ishmael. Some years ago- never mind how long precisely- having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen and regulating the circulation. Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people’s hats off- then, I account it high time to get to sea as soon as I can. This is my substitute for pistol and ball. With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship. There is nothing surprising in this. If they but knew it, almost all men in their degree, some time or other, cherish very nearly the same feelings towards the ocean with me.”*

—*Moby Dick*, 1851.

## 1 Método

Primeiramente, foi criado um arquivo com o nome “moby-dick” usando o comando:

```
cat > moby-dick
```

Neste arquivo foi colocado o texto que foi usado para fazer a análise estatística das palavras, usando comandos como *tr*, *sort*, *uniq* e *awk*. A sequência de comandos seguidos foi:

Para retirar todos os números:

```
tr -d [:digit:] < moby-dick > moby-dick2
```

Para retirar todas as pontuações:

```
tr -d [:punct:] < moby-dick2 > moby-dick3
```

Para mudar letras maiúsculas para minúsculas:

```
tr [:upper:] [:lower:] < moby-dick3 > moby-dick4
```

Para transformar todos os espaços em nova linha:

```
tr -s [:space:] "\n" < moby-dick4 > moby-dick5
```

Para colocar as palavras em ordem alfabética:

```
sort moby-dick5 > moby-dick6
```

Para mostrar as palavras apenas uma vez mas com a frequência que aparecem:

```
uniq -c moby-dick6 > moby-dick7
```

Para colocar em ordem de frequência:

```
sort -n moby-dick7 > moby-dick8
```

Para ver quantas vezes aparece cada palavra (mais a menos frequente) e salvar em *.dat* para plotagem do gráfico:

```
sort -nr moby-dick7 | awk '{print $1}' | uniq | awk '{print NR, $1}' > moby-dick_freqrank.dat
```

Para ver quantas palavras aparecem quantas vezes (menos a mais frequente) e salvar em *.dat* para plotagem do gráfico:

```
awk '{print $1}' moby-dick7 | sort -n | uniq -c > moby-dick_wordfreq.dat
```

Para contar número total de palavras:

```
awk '{ sum += $1 } END {print sum}' moby-dick8 > moby-dick_total
```

Para contar número de palavras diferentes:

```
awk 'END {print NR}' moby-dick8 > moby-dick_diffwords
```

Ver quantas vezes aparece a palavra mais frequente:

```
awk 'END {print $1}' moby-dick8 > moby-dick_morefreq
```

Ver quantas palavras aparecem uma só vez:

```
grep "1 " moby-dick8 | awk '{sum += $1} END {print sum}' > moby-dick_onetime
```

Então foram plotados os gráficos de *Frequência x Ranking* e *# Palavras x Frequência*. Para isso foi usado o *gnuplot*. Dentro do *gnuplot*, os comandos seguidos para a plotagem de *Frequência x Ranking* foram:

```
plot "moby-dick_freqrank.dat" u 1:2 w boxes title "Frequencia x Ranking" lc 3
set style fill pattern 2 border
set style rectangle back fc lt -3 fillstyle solid 1.00 border lt -1
set xlabel "Ranking"
set ylabel "Frequencia"
set boxwidth 0.75
replot
set term pos enh color "Times" 18
set out "moby-dick_freqrank.eps"
replot
unset out
set term wxt
```

Dessa forma, depois de configurado, o gráfico é salvo em formato gráfico portátil (*.eps*).

Da mesma forma, os comandos seguidos para a plotagem, configuração e salvamento em formato gráfico portátil de *# Palavras x Frequência* foram:

```
plot "moby-dick_wordfreq.dat" u 2:1 w boxes title "# Palavras x Frequencia" lc 3
set style fill pattern 2 border
set style rectangle back fc lt -3 fillstyle solid 1.00 border lt -1
set xlabel "Frequencia"
set ylabel "# Palavras"
set xrange [1:1000]
set log
set boxwidth 0.01
set yrange [0.5:2500]
replot
```

```

set term pos enh color "Times" 18
set out "moby-dick_wordfreq.eps"
replot
unset out
set term wxt

```

## 2 Resultados

Nesta seção estão relatados os resultados obtidos.

Estadística	Número
Total de palavras	11467
Palavras diferentes	3139
Veze que a palavra mais frequente aparece	777
Palavras que aparecem apenas uma vez	2449

Tabela 1: Tabela com estatísticas

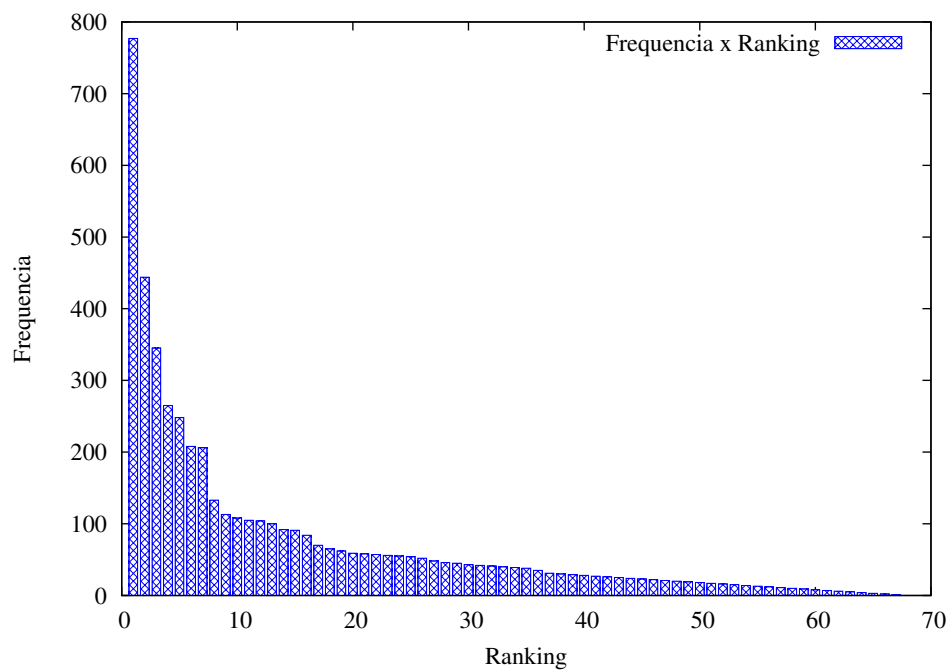


Figura 1: Gráfico mostrando quantas vezes aparece cada palavra, da mais frequente a menos.

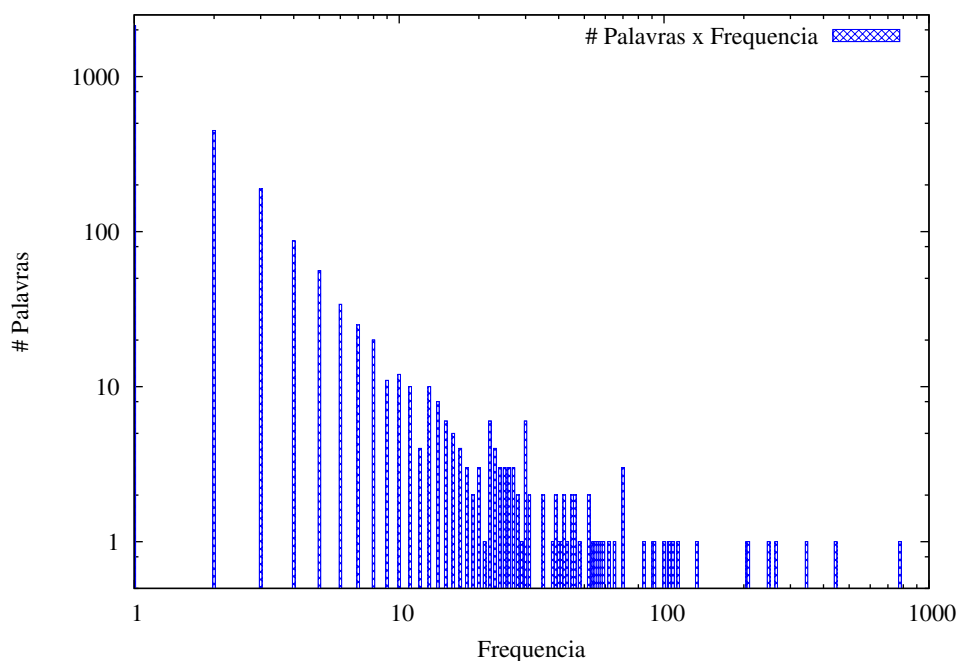


Figura 2: Gráfico mostrando quantas palavras aparecem quantas vezes, da menos frequente a mais, em escala logarítmica.

### 3 Conclusões

Uma estimativa empírica para a curva dos resultados do gráficos de *Frequência x Rankink* poderia ser dada por  $y = \alpha \frac{1}{x} + \beta$ . Para o gráfico de *# Palavras x Frequência*, uma estimativa empírica para a curva, poderia ser dada por  $y = \alpha e^{-\beta x} + \gamma$ . Se o gráfico não estivesse em escala logarítmica, a estimativa para a curva desse gráfico também seria  $y = \alpha \frac{1}{x} + \beta$ .

### Referências

- [1] H. Melville, *Moby-Dick; or, The Whale - Chapter 1: Loomings*. Disponível em: <<https://americanliterature.com/author/herman-melville/book/moby-dick-or-the-whale/chapter-1-loomings>>. Acesso em 28 mar. 2016.
- [2] H. Melville, *Moby-Dick; or, The Whale - Chapter 41: Moby Dick*. Disponível em: <<https://americanliterature.com/author/herman-melville/book/moby-dick-or-the-whale/chapter-41-moby-dick>>.

melville/book/moby-dick-or-the-whale/chapter-41-moby-dick>. Acesso em 28 mar. 2016.

- [3] H. Melville, *Moby-Dick; or, The Whale - Chapter 42: The Whiteness of The Whale*. Disponível em: <<https://americanliterature.com/author/herman-melville/book/moby-dick-or-the-whale/chapter-42-the-whiteness-of-the-whale>>. Acesso em 28 mar. 2016.
- [4] H. Melville, *Moby-Dick; or, The Whale - Chapter 132: The Symphony*. Disponível em: <<https://americanliterature.com/author/herman-melville/book/moby-dick-or-the-whale/chapter-132-the-symphony>>. Acesso em 28 mar. 2016.
- [5] H. Melville, *Moby-Dick; or, The Whale - Epilogue*. Disponível em: <<https://americanliterature.com/author/herman-melville/book/moby-dick-or-the-whale/epilogue>>. Acesso em 28 mar. 2016.