

DSA com a Bíblia



Análise exploratória de dados bíblicos através da linguagem R

AUTOR

Guilherme Bigois

DATA DE PUBLICAÇÃO

13 de fevereiro de 2025

RESUMO

Este trabalho apresenta uma análise exploratória de dados bíblicos utilizando a linguagem R, com o objetivo de identificar padrões e insights relevantes sobre os textos sagrados.

PALAVRAS-CHAVE

Bíblia, R, análise de dados, DSA, exploração de dados

1 Introdução

Nesta seção descreverei o motivo de criação do projeto, os recursos utilizados e os pré-requisitos técnicos para execução do *script* em ambiente local.

1.1 Links úteis

- Link do repositório: <https://github.com/bigois/dsa-bible>
- Meu GitHub: <https://github.com/bigois>
- Meu LinkedIn: <https://www.linkedin.com/in/bigois>

1.2 Objetivo

Após a conclusão do curso [R para Data Science: Conhecendo a linguagem](#) da [Alura](#), acredito ser interessante aplicar os conhecimentos obtidos em um projeto completamente apartado.

De maneira sucinta, neste curso aprendemos:

- Os principais comandos e estruturas de dados do R;
- Como utilizar [estruturas condicionais e de repetição](#);
- [Análise exploratória](#) para responder perguntas sobre os dados;
- Criação e manipulação de [vetores](#), [matrizes](#) e [fatores](#) no R.

Desta forma, tentarei replicar todo o conhecimento aprendido nesta análise de base de dados SQLite.

1.3 Por que a Bíblia Sagrada?

Sempre achei interessantes curiosidades bíblicas, tais como, por exemplo, a **quantidade de capítulos dos livros de Provérbios e Eclesiastes**, os quais têm respectivamente **31 e 12 capítulos**.

Olhando assim, até parece que eles foram projetados para serem lidos **um dia por mês** (no caso de Provérbios) e **um mês por ano** (em se tratando de Eclesiastes).

Logo, escolhi uma base de dados que contivesse a Bíblia Sagrada em forma fácil de ser acessada via SQL, contudo essa base só foi possível de ser obtida graças ao usuário do GitHub [damarals \(Daniel\)](#)

[Amaral](#)), o qual em seu projeto, disponibiliza diversas traduções da Bíblia nos formatos XML, JSON e SQLite.

Eu escolhi a versão [Almeida Revisada e Corrigida \(ARC\)](#), devido à sua familiaridade com a minha primeira Bíblia, todavia para acessar outras traduções, acesse este [link](#).

1.4 Pré-requisitos

Nesta seção listo os arquivos e binários necessários para a execução bem-sucedida do programa.

1.4.1 Base de dados

Para execução deste *script* será necessária a existência de uma das traduções da Bíblia em formato .sqlite contidas no projeto acima citado.

As primeiras linhas do *script* irão perguntar pelo caminho do arquivo, porém é possível fixar o caminho diretamente no código através da variável `path`:

```
1 # path <- file.choose()
2
3 # Caso prefira, fixe o caminho para agilizar a execução do script
4 path <- "C:\\Users\\bigois\\Documents\\Projetos\\dsa-bible\\data\\ARC.sqlite"
```

1.4.2 Versão do R

Para este projeto, foi utilizada a versão [4.3.3 \(2024-02-29 ucrt\)](#) da linguagem R:

```
1 R.version
```

```
platform      _
arch           x86_64-w64-mingw32
os            mingw32
crt            ucrt
system        x86_64, mingw32
status
major         4
minor         3.3
year          2024
month         02
day           29
svn rev       86002
language      R
version.string R version 4.3.3 (2024-02-29 ucrt)
nickname      Angel Food Cake
```

2 Carga de dados

Nesta seção, iremos realizar a extração dos dados necessários para nossa análise.

2.1 Instalação de bibliotecas

Para execução deste código, precisaremos das seguintes bibliotecas:

- [RSQLite](#): Conexão em base de dados SQLite;
- [glue](#): Interpolação de cadeia de caracteres (*string*);
- [dplyr](#): Manipulação de transformação de dados.

```
1 libs <- c("RSQLite", "glue", "dplyr")
2 for (lib in libs) {
3   if (!require(lib, quietly = TRUE, character.only = TRUE)) {
4     install.packages(lib)
5     library(lib, character.only = TRUE)
6   }
7 }
```

2.2 Conexão com a base de dados

Para realizar a conexão com a base de dados, utilizaremos as funções `dbConnect` e `SQLite()` para definir o driver e o objeto de conexão:

```
1 con <- dbConnect(SQLite(), path)
2 tables <- dbListTables(con)
3 tables
```

```
[1] "book"      "metadata"  "testament" "verse"
```

2.3 Obtendo dados

Inicialmente vamos carregar todos os dados de forma completa em objetos da linguagem R, contudo é importante enfatizar **que muitos dos relacionamentos que faremos através do R, poderiam ser facilmente tratados com comandos SQL**, tais como `JOIN`, `MAX`, `MIN` e `AVG`.

Neste ponto, criaremos um objeto para cada tabela através de criação e atribuição dinâmica de variáveis por meio do uso da função `assign`, contudo para definir as variáveis e criar as consultas SQL, utilizaremos as funções `glue_sql` e `glue` do pacote `glue`:

```
1 for (table in tables) {
2   data <- dbGetQuery(con, glue_sql("SELECT * FROM `{table}`", .con = con))
3   assign(glue("db_{table}"), data)
4 }
5
6 head(db_metadata, 3)
```

	key	value
1	version	1
2	name	ARC
3	copyright	Almeida Revista e Corrigida - 1995

```
1 head(db_testament, 3)
```

	id	name
1	1	Antigo Testamento

```
1 head(db_book, 3)
```

	id	book_reference_id	testament_reference_id	name
1	1	1	1	Gênesis
2	2	2	1	Êxodo
3	3	3	1	Levítico

```
1 head(db_verse, 3)
```

	id	book_id	chapter	verse
1	1	1	1	1
2	2	1	1	2
3	3	1	1	3

text

1	No
princípio, criou Deus os céus e a terra.	
2	E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e o Espírito de Deus se movia sobre a face das águas.
3	E
disse Deus: Haja luz. E houve luz.	

2.4 Fechando conexão

Uma vez que os dados necessários já foram obtidos, é uma **boa prática** fechar a conexão com a base de dados:

```
1 dbDisconnect(con)
```

3 Tratamento de dados

Dentre as etapas de um [ETL \(Extract, Transform, Load\)](#), a seção anterior se encaixaria como letra “E” (*extract*), assim como esta seção teria a letra “T” (*transform*) como bem apropriada.

Isto posto, iremos manipular os *data frames* criados de forma a prepará-los para a nossa terceira parte, a análise dos dados:

3.1 Tabela com relacionamentos

Neste ponto, criaremos um objeto chamado `db_full` contendo apenas os dados macro de cada livro da Bíblia, sendo eles:

- `testament_name`: Título de parte da Bíblia;
- `book_name`: Nome do livro;
- `total_chapters`: Total de capítulos contidos no livro;
- `total_verses`: Total de versículos do livro como um todo;
- `total_characters`: Total de caracteres por versículo;

Contudo para que isto seja possível e fácil de realizar, utilizaremos a função `inner_join` da biblioteca `dplyr`:

```
1 db_full <- db_testament %>%
2   inner_join(db_book, by = c("id" = "testament_reference_id")) %>%
3   inner_join(db_verse, by = c("book_reference_id" = "book_id"))
4
5 head(db_full, 3)
```

	id.x	name.x	id.y	book_reference_id	name.y	id.y.y	chapter	verse
1	1	Antigo Testamento	1	1	Gênesis	1	1	1
2	1	Antigo Testamento	1	1	Gênesis	2	1	2
3	1	Antigo Testamento	1	1	Gênesis	3	1	3

text

1		No
princípio, criou Deus os céus e a terra.		
2	E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e o Espírito de Deus se movia sobre a face das águas.	
3		E
disse Deus: Haja luz. E houve luz.		

3.2 Renomeando dados

Como pode ser visto no resultado do *chunk* anterior, após o relacionamento das tabelas, as colunas `id` e `name` passaram a se chamar `id.x`, `id.y`, `name.x` e `name.y`, isto posto, iremos alterar o nome dessas colunas por meio da função `rename`, também do pacote `dplyr`:

```
1 db_full <- db_full %>%
2   rename(testament_name = name.x) %>%
3   rename(book_name = name.y)
4
5 head(db_full, 3)
```

	id.x	testament_name	id.y	book_reference_id	book_name	id.y.y	chapter	verse
1	1	Antigo Testamento	1	1	Gênesis	1	1	1
2	1	Antigo Testamento	1	1	Gênesis	2	1	2
3	1	Antigo Testamento	1	1	Gênesis	3	1	3

text

1		No
princípio, criou Deus os céus e a terra.		
2	E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e o Espírito de Deus se movia sobre a face das águas.	
3		E
disse Deus: Haja luz. E houve luz.		

3.3 Criando novas colunas

Atualmente temos a coluna `text`, a qual contém o texto propriamente dito de cada versículo, contudo para nossa análise, iremos calcular **apenas a quantidade de caracteres contidos em cada versículo** por meio das funções `mutate` (`dplyr`) e `nchar` (`base`):

```

1 db_full <- db_full %>%
2   mutate(total_characters = nchar(text))
3
4 head(db_full, 3)

```

	id.x	testament_name	id.y	book_reference_id	book_name	id.y.y	chapter	verse
1	1	Antigo Testamento	1	1	Gênesis	1	1	1
2	1	Antigo Testamento	1	1	Gênesis	2	1	2
3	1	Antigo Testamento	1	1	Gênesis	3	1	3

text

1		No
princípio, criou Deus os céus e a terra.		
2	E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e o Espírito de Deus se movia sobre a face das águas.	
3		E
disse Deus: Haja luz. E houve luz.		
	total_characters	
1	43	
2	125	
3	36	

3.4 Removendo colunas

2/4 do nosso objetivo para esta seção já estão concluídos, pois criamos as colunas `testament_name` e `book_name`, agora só falta criar as colunas `total_chapters` e `total_verses`.

Contudo, temos algumas colunas como `id.x`, `id.y`, `id.y.y` e `book_reference_id` que por momento não iremos utilizar.

Assim sendo, removeremos essas colunas com o uso da função `select` também contida no pacote `dplyr` em conjunto com o operador `-`:

```

1 db_full <- db_full %>%
2   select(-id.x, -id.y, -id.y.y, -book_reference_id)
3
4 head(db_full, 3)

```

	testament_name	book_name	chapter	verse
1	Antigo Testamento	Gênesis	1	1
2	Antigo Testamento	Gênesis	1	2
3	Antigo Testamento	Gênesis	1	3

text

1		No
princípio, criou Deus os céus e a terra.		
2	E a terra era sem forma e vazia; e havia trevas sobre a face do abismo; e o Espírito de Deus se movia sobre a face das águas.	
3		E
disse Deus: Haja luz. E houve luz.		
	total_characters	
1	43	

```
2          125
3          36
```

3.5 Agrupamento e contagem

Agora precisamos **contar os versículos e caracteres por capítulo** e para isso utilizaremos as funções `group_by` e `summarize` e criaremos um objeto chamado `db_by_chapter`:

```
1 db_by_chapter <- db_full %>%
2   select(-text) %>%
3   group_by(testament_name, book_name, chapter) %>%
4   summarize(
5     total_verses = n(),
6     total_characters = sum(total_characters)
7   )
8
9 head(db_by_chapter, 3)
```

```
# A tibble: 3 × 5
# Groups:   testament_name, book_name [1]
  testament_name book_name chapter total_verses total_characters
  <chr>          <chr>      <int>      <int>      <int>
1 Antigo Testamento 1 Crônicas      1         54         3608
2 Antigo Testamento 1 Crônicas      2         55         4310
3 Antigo Testamento 1 Crônicas      3         24         1975
```

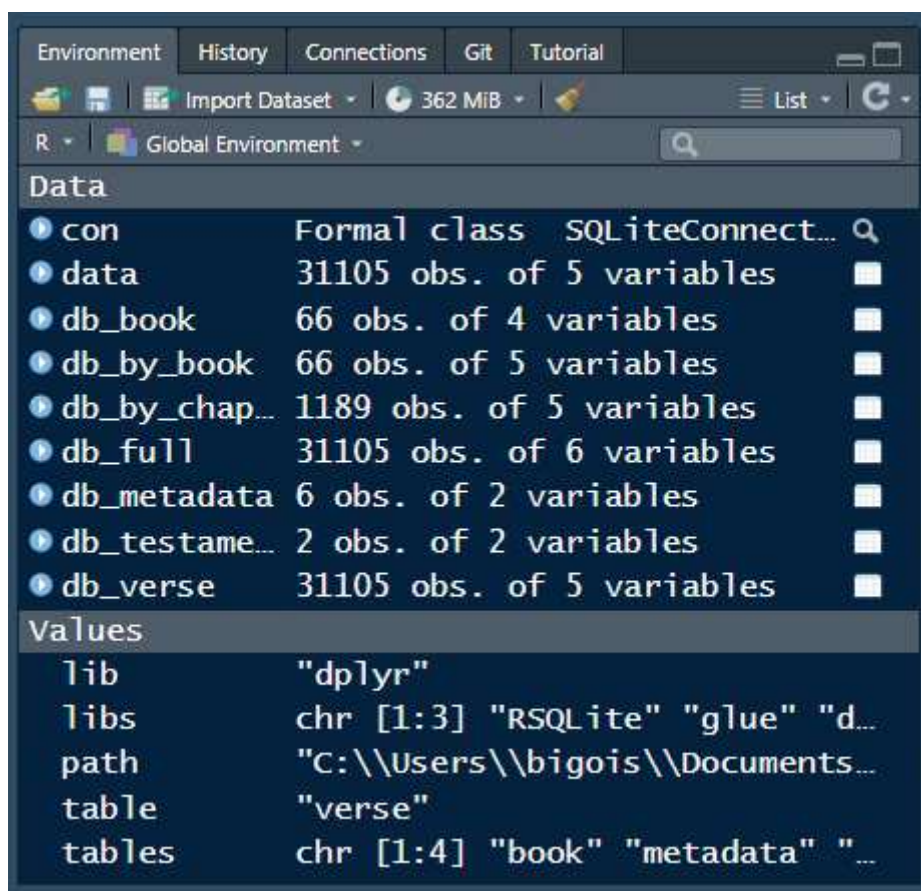
Uma vez contados os versículos e caracteres por capítulo, agora iremos **somar os capítulos por livro** e atribuí-los ao objeto `db_by_book`:

```
1 db_by_book <- db_by_chapter %>%
2   group_by(testament_name, book_name) %>%
3   summarize(
4     total_chapters = n(),
5     total_verses = sum(total_verses),
6     total_characters = sum(total_characters)
7   )
8
9 head(db_by_book, 3)
```

```
# A tibble: 3 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>      <int>      <int>      <int>
1 Antigo Testamento 1 Crônicas      29         942        102465
2 Antigo Testamento 1 Reis          22         817        116364
3 Antigo Testamento 1 Samuel         31         811        120677
```

3.6 Remoção de objetos

Atualmente o painel *Environment* do nosso R Studio encontra-se cheio de variáveis sem uso:

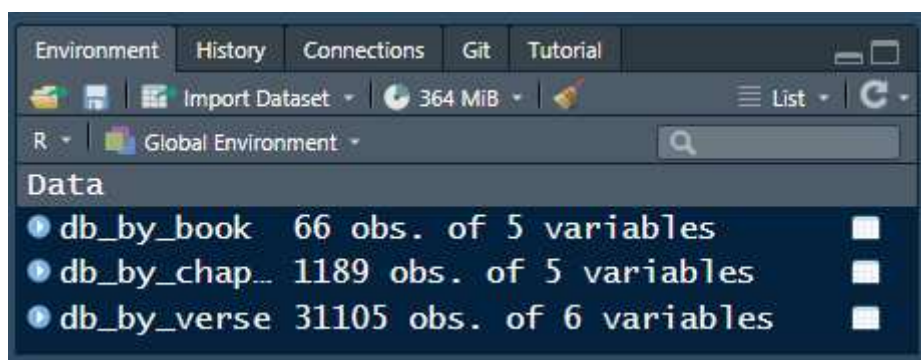


Painel Environment do RStudio com muitas variáveis

Para mantermos a organização dos nossos objetos, iremos remover aqueles aos quais não mais utilizamos através da função `rm`:

```
1 db_by_verse <- db_full
2 rm(con, data, db_full, db_metadata, db_testament, db_book, db_verse, lib, libs, path, table, tables)
```

Desta forma, ficamos com o painel Environment desta forma:



Painel Environment do RStudio com poucas variáveis

4 Análise de dados

Na última etapa deste projeto, validaremos e analisaremos as informações contidas nos objetos `db_by_book`, `db_by_chapter` e `db_by_verse`.

Toda análise é baseada em **perguntas e constatações por meio de observação**, portanto, responderemos a algumas perguntas formuladas para compreender melhor a base de dados analisada.

4.1 Maiores e menores

Para descobrirmos qual é o maior e o menor livro da Bíblia, devemos entender que **a relação de maior/menor se dá com base no ponto de vista abordado**, pois podemos definir o que é “ser maior” ou “ser menor” de **forma quantitativa com base na quantidade de caracteres, versículos ou capítulos**.

Dito isso, definiremos o tamanho de cada livro com base nos três atributos mencionados acima, utilizando as seguintes colunas:

- `total_characters` do objeto `db_by_verse`;
- `total_verses` do objeto `db_by_chapter`;
- `total_chapters` do objeto `db_by_book`.

4.1.1 Maior versículo com base na quantidade de caracteres

```
1 bigger_verse_by_characters <- which.max(db_by_verse$total_characters)
2 db_by_verse[bigger_verse_by_characters, ]
```

	testament_name	book_name	chapter	verse
12830	Antigo Testamento	Ester	8	9

text

12830 Então, foram chamados os escrivães do rei, naquele mesmo tempo e no mês terceiro (que é o mês de sivã), aos vinte e três do mesmo, e se escreveu conforme tudo quanto ordenou Mardoqueu aos judeus, como também aos sátrapas, e aos governadores, e aos maioraes das províncias que se estendem da Índia até à Etiópia, cento e vinte e sete províncias, a cada província segundo a sua escritura e a cada povo conforme a sua língua; como também aos judeus segundo a sua escritura e conforme a sua língua.

	total_characters
12830	494

4.1.2 Menor versículo com base na quantidade de caracteres

```
1 smaller_verse_by_characters <- which.min(db_by_verse$total_characters)
2 db_by_verse[smaller_verse_by_characters, ]
```

	testament_name	book_name	chapter	verse	text	total_characters
25813	Novo Testamento	Lucas	20	30	e o segundo	11

4.1.3 Maior capítulo com base na quantidade de caracteres

```
1 bigger_chapter_by_characters <- which.max(db_by_chapter$total_characters)
2 db_by_chapter[bigger_chapter_by_characters, ]
```

A tibble: 1 × 5

Groups: testament_name, book_name [1]

testament_name	book_name	chapter	total_verses	total_characters
----------------	-----------	---------	--------------	------------------

	<chr>	<chr>	<int>	<int>	<int>
1	Antigo Testamento	Salmos	119	176	12986

4.1.4 Menor capítulo com base na quantidade de caracteres

```
1 smaller_chapter_by_characters <- which.min(db_by_chapter$total_characters)
2 db_by_chapter[smaller_chapter_by_characters, ]
```

```
# A tibble: 1 × 5
# Groups:   testament_name, book_name [1]
  testament_name book_name chapter total_verses total_characters
  <chr>          <chr>    <int>      <int>          <int>
1 Antigo Testamento Salmos      117         2             162
```

4.1.5 Maior capítulo com base na quantidade de versículos

```
1 bigger_chapter_by_verses <- which.max(db_by_chapter$total_verses)
2 db_by_chapter[bigger_chapter_by_verses, ]
```

```
# A tibble: 1 × 5
# Groups:   testament_name, book_name [1]
  testament_name book_name chapter total_verses total_characters
  <chr>          <chr>    <int>      <int>          <int>
1 Antigo Testamento Salmos      119      176          12986
```

4.1.6 Menor capítulo com base na quantidade de versículos

```
1 smaller_chapter_by_verses <- which.min(db_by_chapter$total_verses)
2 db_by_chapter[smaller_chapter_by_verses, ]
```

```
# A tibble: 1 × 5
# Groups:   testament_name, book_name [1]
  testament_name book_name chapter total_verses total_characters
  <chr>          <chr>    <int>      <int>          <int>
1 Antigo Testamento Salmos      117         2             162
```

4.1.7 Maior livro com base na quantidade de caracteres

```
1 bigger_book_by_characters <- which.max(db_by_book$total_characters)
2 db_by_book[bigger_book_by_characters, ]
```

```
# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>          <int>      <int>          <int>
1 Antigo Testamento Salmos           150      2461          210005
```

4.1.8 Menor livro com base na quantidade de caracteres

```

1  smaller_book_by_characters <- which.min(db_by_book$total_characters)
2  db_by_book[smaller_book_by_characters, ]

```

```

# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>          <int>         <int>         <int>
1 Novo Testamento 3 João              1             15             1411

```

4.1.9 Maior livro com base na quantidade de versículos

```

1  bigger_book_by_verses <- which.max(db_by_book$total_verses)
2  db_by_book[bigger_book_by_verses, ]

```

```

# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>          <int>         <int>         <int>
1 Antigo Testamento Salmos             150         2461         210005

```

4.1.10 Menor livro com base na quantidade de versículos

```

1  smaller_book_by_verses <- which.min(db_by_book$total_verses)
2  db_by_book[smaller_book_by_verses, ]

```

```

# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>          <int>         <int>         <int>
1 Novo Testamento 2 João              1             13             1469

```

4.1.11 Maior livro com base na quantidade de capítulos

```

1  bigger_book_by_chapters <- which.max(db_by_book$total_chapters)
2  db_by_book[bigger_book_by_chapters, ]

```

```

# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters
  <chr>          <chr>          <int>         <int>         <int>
1 Antigo Testamento Salmos             150         2461         210005

```

4.1.12 Menor livro com base na quantidade de capítulos

```

1  smaller_book_by_chapters <- which.min(db_by_book$total_chapters)
2  db_by_book[smaller_book_by_chapters, ]

```

```

# A tibble: 1 × 5
# Groups:   testament_name [1]
  testament_name book_name total_chapters total_verses total_characters

```

<chr>	<chr>	<int>	<int>	<int>
1 Antigo Testamento	Obadias	1	21	3054

4.1.13 Compilando resultados

Agora que capturamos os dados necessários sobre os maior e menores livro da Bíblia, vamos juntá-los em uma matriz que nos permita uma melhor visualização, e será necessário o uso das funções `c` e `matrix`:

```
1 bigger_book_names <- c(
2   db_by_book[bigger_book_by_characters, 'book_name'],
3   db_by_book[bigger_book_by_verses, 'book_name'],
4   db_by_book[bigger_book_by_chapters, 'book_name']
5 )
6
7 bigger_book_totals <- c(
8   db_by_book[bigger_book_by_characters, 'total_characters'],
9   db_by_book[bigger_book_by_verses, 'total_verses'],
10  db_by_book[bigger_book_by_chapters, 'total_chapters']
11 )
12
13 smaller_book_names <- c(
14   db_by_book[smaller_book_by_characters, 'book_name'],
15   db_by_book[smaller_book_by_verses, 'book_name'],
16   db_by_book[smaller_book_by_chapters, 'book_name']
17 )
18
19 smaller_book_totals <- c(
20   db_by_book[smaller_book_by_characters, 'total_characters'],
21   db_by_book[smaller_book_by_verses, 'total_verses'],
22   db_by_book[smaller_book_by_chapters, 'total_chapters']
23 )
24
25 bigger_books <- matrix(c(bigger_book_names, bigger_book_totals), ncol = 2, nrow = 3)
26 smaller_books <- matrix(c(smaller_book_names, smaller_book_totals), ncol = 2, nrow = 3)
```

Agora vamos nomear as linhas e colunas das matrizes criadas de forma que nos permita uma melhor compreensão dos dados compilados, e para isto usaremos as funções `rownames` e `colnames`:

```
1 size_rownames <- c("Caracteres", "Versículos", "Capítulos")
2 size_colnames <- c("Livro", "Total")
3
4 rownames(bigger_books) <- size_rownames
5 rownames(smaller_books) <- size_rownames
6
7 colnames(bigger_books) <- size_colnames
8 colnames(smaller_books) <- size_colnames
9
10 rm(bigger_book_names, bigger_book_totals, smaller_book_names, smaller_book_totals)
```

Por fim teremos os maiores livro da Bíblia separados por categoria:

```
1 # Maiores livros com base na quantidade de caracteres, versículos e capítulos
2 bigger_books
```

	Livro	Total
Caracteres	"Salmos"	210005
Versículos	"Salmos"	2461
Capítulos	"Salmos"	150

Assim como também os menores:

```
1 # Menores livros com base na quantidade de caracteres, versículos e capítulos
2 smaller_books
```

	Livro	Total
Caracteres	"3 João"	1411
Versículos	"2 João"	13
Capítulos	"Obadias"	1

4.2 Médias e medianas

A análise e comparação dos valores de média e mediana permitem **evitar interpretações equivocadas**, pois diferentemente da média, a **mediana ignora as metades superior e inferior**, descartando assim valores **muito discrepantes e extremos** que estejam influenciando na média.

Para ilustrar a compreensão da importância da análise de medianas, imagine uma pessoa dizendo a frase abaixo:

"Minha empresa é um ótimo lugar para se trabalhar. Lá a média salarial é de 6 mil reais!"

Contudo, para ilustrar a falácia acima apresentada, definimos o seguinte cenário:

```
1 salary_boss <- 20000
2 salary_coordinator <- 10000
3 salary_employee_1 <- 1500
4 salary_employee_2 <- 1500
5 salary_employee_3 <- 1500
6 salary_employee_4 <- 1500
7
8 company_salary <- c(salary_boss, salary_coordinator, salary_employee_1, salary_employee_2, salary_employee_3, salary_employee_4)
9
10 salary_mean <- round(mean(company_salary))
11 salary_median <- round(median(company_salary))
12 salary_diff <- round(abs(salary_mean - salary_median))
13
14 paste(glue('Média: {salary_mean} | Mediana: {salary_median} | Diferença: {salary_diff}'))
```

```
[1] "Média: 6000 | Mediana: 1500 | Diferença: 4500"
```

Viu como números não mentem, mas mentirosos usam números para mentir? 😂

No caso acima a média permite que os salários do chefe (`salary_boss`) e do coordenador (`salary_coordinator`) **influenciem a média final, enviesando a análise** com uma ideia de que a faixa salarial é alta para todos, contudo, **a mediana nos mostra que isso não é verdade**.

Com isso em mente, seguimos com nossa análise baseada na Bíblia...

4.2.1 Caracteres por versículo

```
1 mean_characters_by_verse <- round(mean(db_by_verse[, "total_characters"]))
2 median_characters_by_verse <- round(median(db_by_verse[, "total_characters"]))
3 diff_characters_by_verse <- round(abs(mean_characters_by_verse - median_characters_by_verse))
4
5 paste(glue('Média: {mean_characters_by_verse} | Mediana: {median_characters_by_verse} | Diferença: {diff_characters_by_verse}'))
```

```
[1] "Média: 122 | Mediana: 114 | Diferença: 8"
```

4.2.2 Caracteres por capítulo

```
1 mean_characters_by_chapter <- round(mean(db_by_chapter[["total_characters"]]))
2 median_characters_by_chapter <- round(median(db_by_chapter[["total_characters"]]))
3 diff_characters_by_chapter <- round(abs(mean_characters_by_chapter - median_characters_by_chapter))
4
5 paste(glue('Média: {mean_characters_by_chapter} | Mediana: {median_characters_by_chapter} | Diferença: {diff_characters_by_chapter}'))
```

```
[1] "Média: 3192 | Mediana: 3014 | Diferença: 178"
```

4.2.3 Caracteres por livro

```
1 mean_characters_by_book <- round(mean(db_by_book[["total_characters"]]))
2 median_characters_by_book <- round(median(db_by_book[["total_characters"]]))
3 diff_characters_by_book <- round(abs(mean_characters_by_book - median_characters_by_book))
4
5 paste(glue('Média: {mean_characters_by_book} | Mediana: {median_characters_by_book} | Diferença: {diff_characters_by_book}'))
```

```
[1] "Média: 57504 | Mediana: 28800 | Diferença: 28704"
```

4.2.4 Versículos por capítulo

```
1 mean_verses_by_chapter <- round(mean(db_by_chapter[["total_verses"]]))
2 median_verses_by_chapter <- round(median(db_by_chapter[["total_verses"]]))
3 diff_verses_by_chapter <- round(abs(mean_verses_by_chapter - median_verses_by_chapter))
4
5 paste(glue('Média: {mean_verses_by_chapter} | Mediana: {median_verses_by_chapter} | Diferença: {diff_verses_by_chapter}'))
```

```
[1] "Média: 26 | Mediana: 24 | Diferença: 2"
```

4.2.5 Versículos por livro

```
1 mean_verses_by_book <- round(mean(db_by_book[["total_verses"]]))
2 median_verses_by_book <- round(median(db_by_book[["total_verses"]]))
3 diff_verses_by_book <- round(abs(mean_verses_by_book - median_verses_by_book))
4
5 paste(glue('Média: {mean_verses_by_book} | Mediana: {median_verses_by_book} | Diferença:
```

```
[1] "Média: 471 | Mediana: 216 | Diferença: 255"
```

4.2.6 Capítulos por livro

```
1 mean_chapters_by_book <- round(mean(db_by_book[["total_chapters"]]))
2 median_chapters_by_book <- round(median(db_by_book[["total_chapters"]]))
3 diff_chapters_by_book <- round(abs(mean_chapters_by_book - median_chapters_by_book))
4
5 paste(glue('Média: {mean_chapters_by_book} | Mediana: {median_chapters_by_book} | Diferen
```

```
[1] "Média: 18 | Mediana: 12 | Diferença: 6"
```

4.2.7 Compilando resultados

Assim como fizemos na [seção anterior](#), vamos formatar os dados obtidos de uma a permitir uma melhor visualização:

```
1 mean_median_matriz <- matrix(c(
2   mean_characters_by_verse,
3   median_characters_by_verse,
4   diff_characters_by_verse,
5   mean_characters_by_chapter,
6   median_characters_by_chapter,
7   diff_characters_by_chapter,
8   mean_characters_by_book,
9   median_characters_by_book,
10  diff_characters_by_book,
11  mean_verses_by_chapter ,
12  median_verses_by_chapter ,
13  diff_verses_by_chapter,
14  mean_verses_by_book,
15  median_verses_by_book,
16  diff_verses_by_book,
17  mean_chapters_by_book,
18  median_chapters_by_book,
19  diff_chapters_by_book
20 ), ncol = 3, byrow = TRUE)
21
22 colnames(mean_median_matriz) <- c("Média", "Mediana", "Diferença")
23 rownames(mean_median_matriz) <- c(
24   'Caracteres por versículo',
25   'Caracteres por capítulo',
26   'Caracteres por livro',
```

```

27 'Versículos por capítulo',
28 'Versículos por livro',
29 'Capítulos por livro'
30 )
31
32 mean_median_matriz

```

	Média	Mediana	Diferença
Caracteres por versículo	122	114	8
Caracteres por capítulo	3192	3014	178
Caracteres por livro	57504	28800	28704
Versículos por capítulo	26	24	2
Versículos por livro	471	216	255
Capítulos por livro	18	12	6

Agora filtraremos apenas indicadores que tenham o valor intermediário (desvio) entre média e mediana menor que 10%:

```

1 less_than_10 <- mean_median_matriz[, 'Diferença'] / ((mean_median_matriz[, 'Média'] + mea
2 mean_median_matriz[less_than_10, ]

```

	Média	Mediana	Diferença
Caracteres por versículo	122	114	8
Caracteres por capítulo	3192	3014	178
Versículos por capítulo	26	24	2

Esta matriz nos permite inferir que na Bíblia temos um valor intermediário comum para caracteres por versículo (133) e capítulo (3103), assim como versículos por capítulo (25).

4.3 Conclusão

Após a obtenção de tantos dados, conseguimos chegar as seguintes conclusões:

1. O maior livro da Bíblia é o livro de Salmos do Antigo Testamento, tendo 210005 caracteres, 2461 versículos e 150 capítulos;
2. O menor livro da da Bíblia em quantidade de caracteres é o livro de 3 João do Novo Testamento, com um total de 1411 caracteres;
3. O menor livro da da Bíblia em quantidade de versículos é o livro de 2 João do Novo Testamento, com um total de 13 caracteres;
4. O menor livro da da Bíblia em quantidade de capítulos é o livro de Obadias do Antigo Testamento, com um total de 1 caracteres;
5. O menor versículo da Bíblia tem 11 caracteres e fica no livro de Lucas no Novo Testamento;
6. O maior versículo da Bíblia tem 494 caracteres e fica no livro de Ester no Antigo Testamento;
7. A Bíblia tem um total de 66 livros, sendo 27 (41%) do Antigo Testamento e 39 do Novo Testamento (59%).

4.4 Ponto de atenção

Os dados que você leu na seção [Conclusão](#) foram escritos utilizando *inline code*, isto posto, caso você avalie o script do projeto no GitHub, verá que os dados não estão fixados no código, mas sim obtidos através de variáveis capturadas anteriormente:

Conclusão

Após a obtenção de tantos dados, conseguimos chegar as seguintes conclusões:

1. O maior livro da Bíblia é o livro de Salmos do `r db_by_book %>% filter(book_name == "Salmos") %>% pull(testament_name)`, tendo `r db_by_book %>% filter(book_name == "Salmos") %>% pull(total_characters)` caracteres, `r db_by_book %>% filter(book_name == "Salmos") %>% pull(total_verses)` versículos e `r db_by_book %>% filter(book_name == "Salmos") %>% pull(total_chapters)` capítulos;
2. O menor livro da da Bíblia em quantidade de caracteres é o livro de `r db_by_book[smaller_book_by_characters,]$book_name` do `r db_by_book[smaller_book_by_characters,]$testament_name`, com um total de `r db_by_book[smaller_book_by_characters,]$total_characters` caracteres;
3. O menor livro da da Bíblia em quantidade de versículos é o livro de `r db_by_book[smaller_book_by_verses,]$book_name` do `r db_by_book[smaller_book_by_verses,]$testament_name`, com um total de `r db_by_book[smaller_book_by_verses,]$total_verses` caracteres;
4. O menor livro da da Bíblia em quantidade de capítulos é o livro de `r db_by_book[smaller_book_by_chapters,]$book_name` do `r db_by_book[smaller_book_by_chapters,]$testament_name`, com um total de `r db_by_book[smaller_book_by_chapters,]$total_chapters` caracteres;
5. O menor versículo da Bíblia tem `r db_by_verse[smaller_verse_by_characters,]$total_characters` caracteres e fica no livro de `r db_by_verse[smaller_verse_by_characters,]$book_name` no `r db_by_verse[smaller_verse_by_characters,]$testament_name`;
6. O maior versículo da Bíblia tem `r db_by_verse[bigger_verse_by_characters,]$total_characters` caracteres e fica no livro de `r db_by_verse[bigger_verse_by_characters,]$book_name` no `r db_by_verse[bigger_verse_by_characters,]$testament_name`;
7. A Bíblia tem um total de `r db_by_book %>% nrow()` livros, sendo `r db_by_book %>% filter(testament_name == 'Novo Testamento') %>% nrow()` (`r round(db_by_book %>% filter(testament_name == 'Novo Testamento') %>% nrow() / db_by_book %>% nrow() * 100)%`) do Antigo Testamento e `r db_by_book %>% filter(testament_name == 'Antigo Testamento') %>% nrow()` do Novo Testamento (`r round(db_by_book %>% filter(testament_name == 'Antigo Testamento') %>% nrow() / db_by_book %>% nrow() * 100)%`).

Inline-code para código incorporado ao texto

4.5 Observações finais

Espero que este projeto possa ser útil tanto para aqueles que desejam explorar dados ou aprender mais sobre a Bíblia, quanto para quem busca aprimorar seus conhecimentos em R.