Primeiro Relatorio em DS4A

Gabriel Miranda

April 04, 2018

Imports

```
library(jsonlite)
library(readxl)
library(readr)
# library(dplyr)
```

Reading the data

```
orient_prof_unb_json <- fromJSON("orientacoes_professores_unb.json")
dir_gru_pesq_unb_xls <- read_excel("diretorio_grupos_pesquisa_unb.xls")
tes_diss_unb_json <- fromJSON("teses_dissertacoes_unb.json")
pub_rep_oasis_txt <- read_tsv("publicacoes_repositorio_oasis.txt")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   oai_identifier_str = col_character(),
##   network_acronym_str = col_character(),
##   network_name_str = col_character(),
##   title = col_character(),
##   author = col_character(),
##   dc.contributor.advisor1.fl_str_mv = col_character(),
##   topic = col_character(),
##   description = col_character(),
##   publishDate = col_character(),
##   format = col_character(),
##   dc.identifier.citation.fl_str_mv = col_character(),
##   url = col_character(),
##   language = col_character(),
##   dc.rights.driver.fl_str_mv = col_character(),
##   repository.mail.fl_str_mv = col_character(),
##   author2 = col_character(),
##   dc.relation.none.fl_str_mv = col_character()
## )
```

```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 10360 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected   actual     file                                                                    exp
## ... .................. ... ....................................................................... ..
## See problems(...) for more details.
```

Showing data structure

```
names(orient_prof_unb_json)
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"
## [2] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"
## [3] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"
```

```
## [4] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"
## [5] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"
## [6] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## [7] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## [8] "ORIENTACAO_CONCLUIDA_MESTRADO"
```

```r
names(dir_gru_pesq_unb_xls)
```

```
## [1] "Sigla Instituição Grupo: UnB" "X__1"
## [3] "X__2"                         "X__3"
## [5] "X__4"                         "X__5"
```

```r
names(tes_diss_unb_json)
```

```
## [1] "responseHeader" "response"
```

```r
names(pub_rep_oasis_txt)
```

```
##  [1] "id"
##  [2] "oai_identifier_str"
##  [3] "network_acronym_str"
##  [4] "network_name_str"
##  [5] "title"
##  [6] "author"
##  [7] "dc.contributor.advisor1.fl_str_mv"
##  [8] "topic"
##  [9] "description"
## [10] "publishDate"
## [11] "format"
## [12] "dc.identifier.citation.fl_str_mv"
## [13] "url"
## [14] "language"
## [15] "dc.rights.driver.fl_str_mv"
## [16] "repository.mail.fl_str_mv"
## [17] "author2"
## [18] "dc.relation.none.fl_str_mv"
```