cleaning-fires.R

jamesmonks

2019-10-15

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v ggplot2 3.2.1.9000     v purrr   0.3.2
## v tibble  2.1.3          v dplyr   0.8.3.9000
## v tidyr   1.0.0          v stringr 1.4.0
## v readr   1.3.1          v forcats 0.4.0
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tsibble)
```

```
##
## Attaching package: 'tsibble'
## The following object is masked from 'package:dplyr':
##
##     id
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:tsibble':
##
##     interval, new_interval
## The following object is masked from 'package:base':
##
##     date
```

```
# Read the unclean data
fires <- read_csv(here::here("data-raw", "amazon.csv"))
```

```
## Warning in readLines(f, n): line 1 appears to contain an embedded nul
## Parsed with column specification:
## cols(
##   year = col_double(),
##   state = col_character(),
##   month = col_character(),
##   number = col_double(),
##   date = col_date(format = "")
## )
```

```
# Fixing the Problem Month
problem_term <- fires %>%
  filter(stringr::str_detect(month, "[^a-zA-Z]")) %>%
  slice(1) %>%
```

```
  pull(month)

fires_date <- fires %>%
  mutate(month_clean = case_when(
    month == "Abril"        ~ "April",
    month == "Agosto"       ~ "August",
    month == "Dezembro"     ~ "December",
    month == "Fevereiro"    ~ "February",
    month == "Janeiro"      ~ "January",
    month == "Julho"        ~ "July",
    month == "Junho"        ~ "June",
    month == "Maio"         ~ "May",
    month == problem_term     ~  "March",
    month == "Novembro"     ~ "November",
    month == "Outubro"      ~ "October",
    month == "Setembro"     ~ "September",
    TRUE                    ~ NA_character_
  ),
  day = 1) %>%
  mutate(date_month = lubridate::ymd(glue::glue("{year}-{month_clean}-{day}"))) %>%
  mutate(date_index = yearmonth(date_month)) %>%
  mutate(state = ifelse(str_detect(state, "Par\xe1"), "Para", state))


# Creating the clean fire object
fires_clean <-
  fires_date %>%
  filter(!(state %in% c("Mato Grosso", "Rio", "Paraiba"))) %>%
  filter(duplicated(.) == FALSE) %>%
  select(-c(month, date, day, date_month)) %>%
  rename(date = date_index, month = month_clean, fires = number) %>%
  mutate(state = ifelse(str_detect(state, "Par\xe1"), "Para", state)) %>%
  as_tsibble(index = date, key = state)



fires_clean_all <-
  fires_date %>%
  filter(!(state %in% c("Mato Grosso", "Rio", "Paraiba"))) %>%
  filter(duplicated(.) == FALSE) %>%
  group_by(date_index) %>%
  summarise(fires = sum(number)) %>%
  as_tsibble(index = date_index) %>%
  rename(date = date_index)


# Save the clean data
write_rds(fires_clean, here::here("data", "fires_clean.rds"))
write_rds(fires_clean_all, here::here("data", "fires_clean_all.rds"))

write_rds(fires_date, here::here("data-raw", "fires_date.rds"))
```