Koronavírus

Granát Marcell és Mazzag Bálint

2021. március 1.

Tartalomjegyzék

Bevezetés **3**

Adatok **3**

Gépi fordítás 3

Leíró statisztikák **6**

Függelék: R kódok **7**

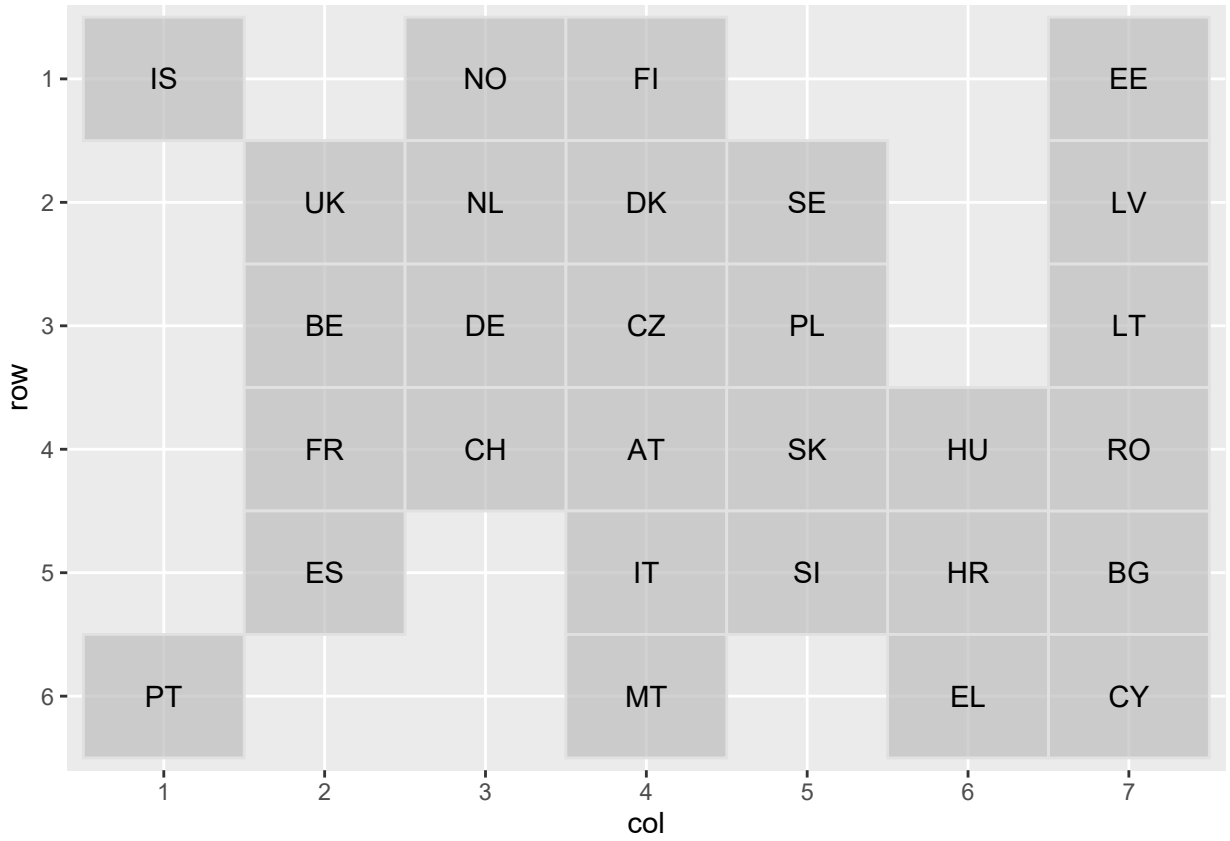
Absztrakt

Here is the abstract.

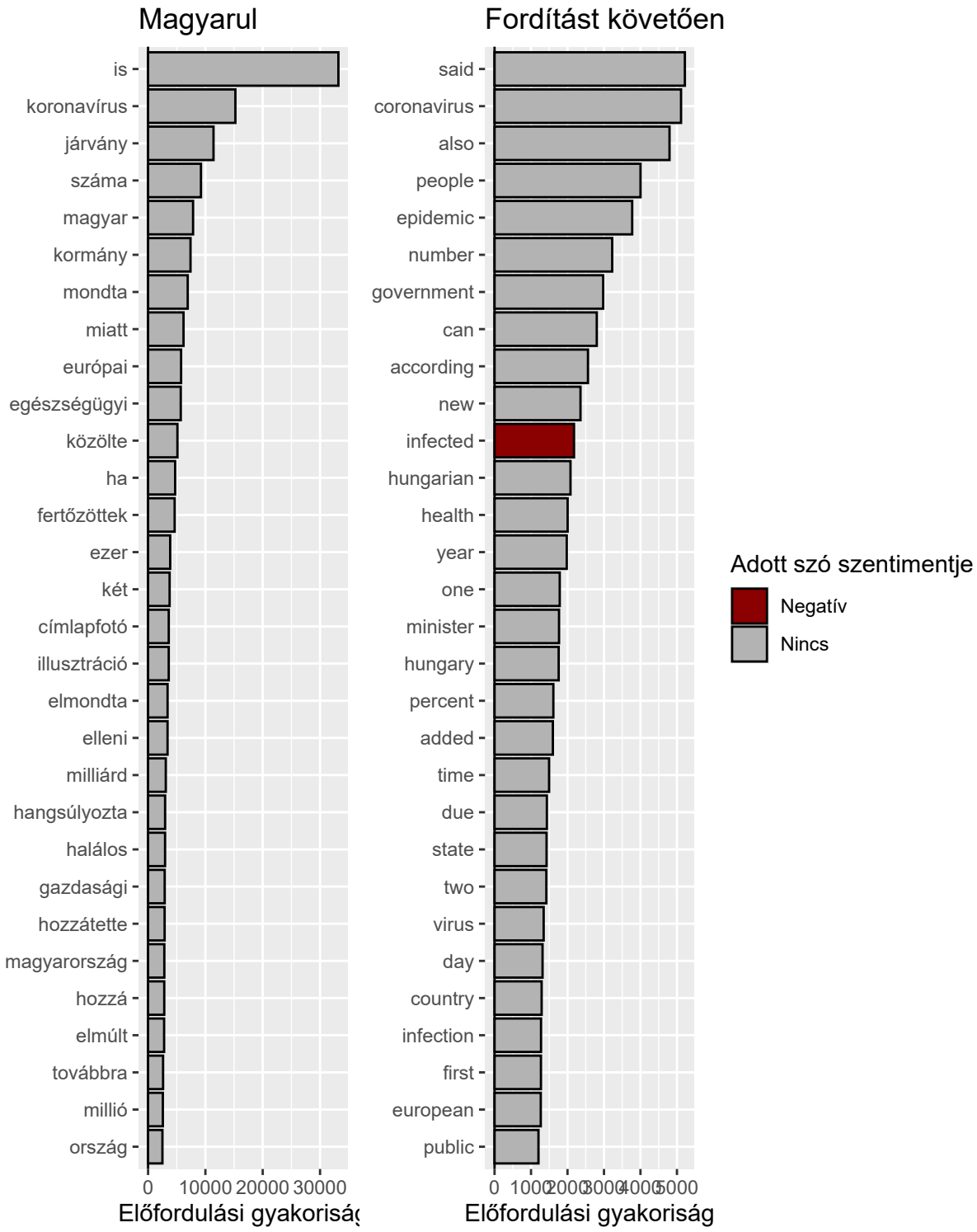
Bevezetés

Adatok

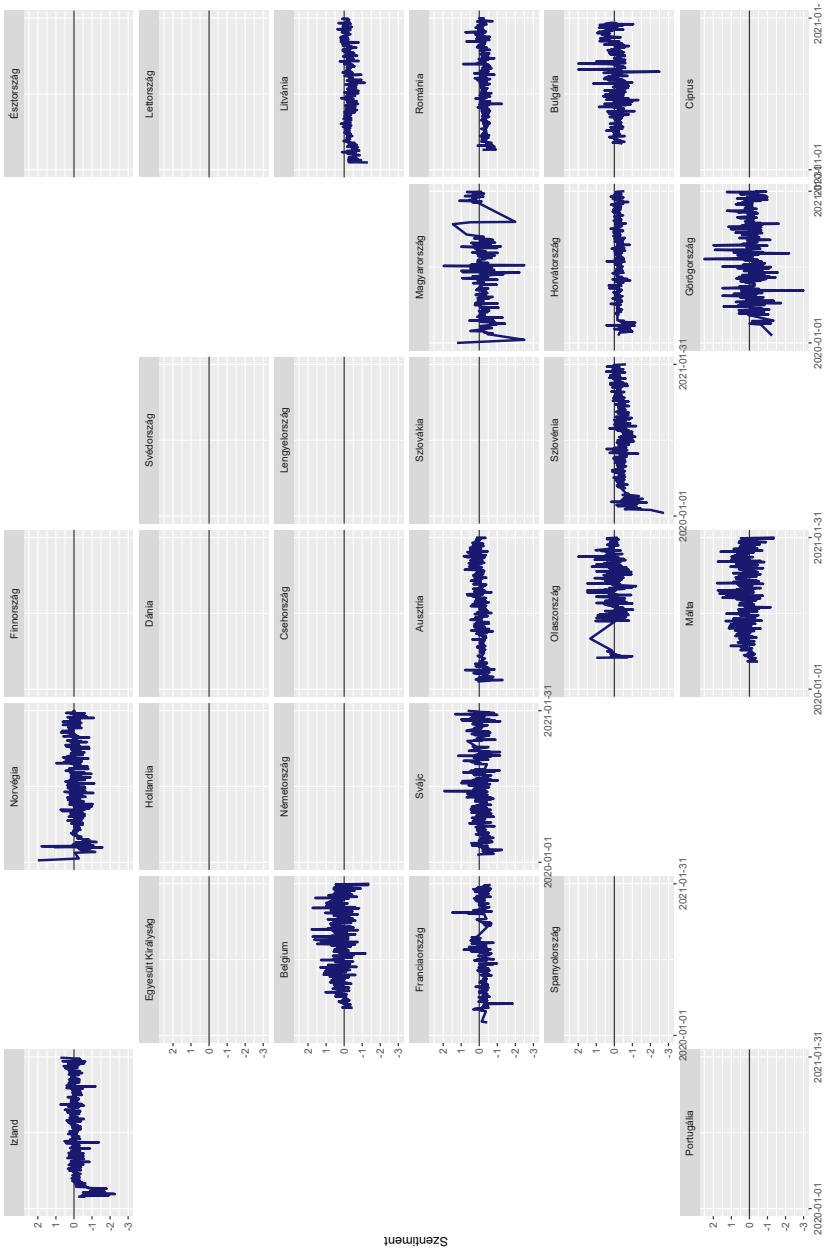
Gépi fordítás




```
Error: Can't subset columns that don't exist.
x Column `value` doesn't exist.
```

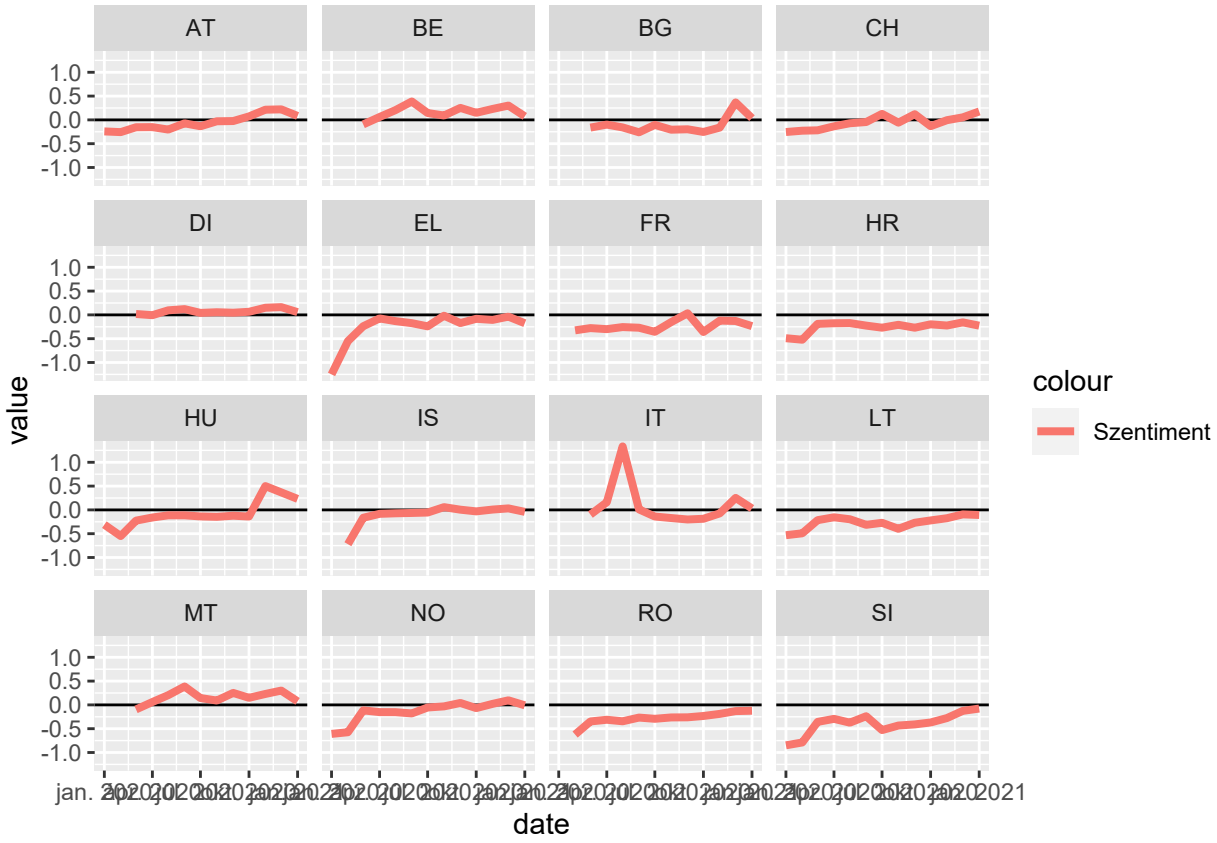


1. ábra. Leggyakrabban előforduló szavak a magyar nylevű cikkekben a fordítást megelőzően és azt követően.



2. ábra. A szentiment alakulása országonként

Leíró statisztikák



Függelék: R kódok

```r
# Set up ------------------------------------------------------------------------------

## Packages ==========================================================================

library(tidyverse)
library(patchwork)
library(knitr)
library(broom)
library(geofacet)
library(tidytext)
library(tm)
library(wordcloud)

## Gg theme ==========================================================================

update_geom_defaults("point", list(fill = "cyan4",
                                   shape = 21,
                                   color = "black",
                                   size = 1.4))
update_geom_defaults("line",
                     list(color = "midnightblue", size = 1.4))

update_geom_defaults("smooth", list(color = "red4", size = 1.4))

update_geom_defaults("density",
                     list(color = "midnightblue", fill =  "midnightblue",
                          alpha = .3, size = 1.4))

extrafont::loadfonts(device="win")

theme_set(theme_grey() + theme(
  legend.direction = "vertical",
  # text = element_text(family = "Impact"),
  plot.caption = element_text(family = "serif")
))

load("dat.RData")
Hungary_rawtext <- readxl::read_excel("scrapping raw csv/Hungary_rawtext.xlsx") %>%
  select(date, title, URL = links, text) %>%
  mutate_all(function(x) str_remove_all(x, "\r")) %>%
  mutate_all(function(x) str_remove_all(x, "\t")) %>%
  mutate_all(function(x) str_remove_all(x, "\n")) %>%
  mutate_at(-1, function(x) zoo::na.locf(x)) %>%
  filter(!str_detect(date, '_x000') & date != '0') %>%
  filter(!str_detect(text, 'mtva_player')) %>% # TODO consider a better solution
  mutate(
    date = gsub(" -.*", "", date),
    text = str_remove_all(text, "_x000D_"),
    date = lubridate::ymd(date)
  ) %>%
  tidytext::unnest_tokens(words, text)

```

```r
dat_sentiment <- dat %>%
  select(date, text, country) %>%
  mutate(country = ifelse(str_detect(country, "BE"), "BE", country)) %>%
  {left_join(tidytext::unnest_tokens(., words, text),
             get_sentiments("afinn"), by=c("words"="word"))}
# TODO other packages

dat_sentiment_daily <- dat_sentiment %>%
  group_by(date, country) %>%
  summarise(value = mean(value, na.rm = T), n = n()) %>%
  ungroup() %>%
  na.omit() %>%
  rename(code = country)

dat_sentiment_monthly <- dat_sentiment %>%
  na.omit() %>%
  mutate(
    date = lubridate::ym(paste(lubridate::year(date), lubridate::month(date), sep = "-"))
  ) %>%
  group_by(date, country) %>%
  summarise(value = mean(value, na.rm = T), n = n()) %>%
  ungroup() %>%
  na.omit() %>%
  rename(code = country)

Hungary_rawtext %>%
  filter(!str_detect(words, '\\d')) %>%
  anti_join(data.frame(words = stopwords::stopwords('hungarian'))) %>%
  count(words, sort = T) %>%
  arrange(desc(n)) %>%
  head(30) %>%
  mutate(
     words = fct_reorder(words, n)
  ) %>%
  ggplot() +
  aes(n, words) +
  geom_vline(xintercept = 0) +
  geom_col(color = 'black', fill = "gray70") +
  labs(title = 'Magyarul', x = 'Előfordulási gyakoriság', y = NULL) +

dat_sentiment %>%
  filter(country == 'HU') %>%
  filter(!str_detect(words, '\\d')) %>%
  anti_join(data.frame(words = stopwords::stopwords())) %>%
  count(words, value, sort = T) %>%
  arrange(desc(n)) %>%
  head(30) %>%
  mutate(
    value = case_when(
      value < 0 ~ "Negatív",
      value > 0 ~ "Pozitív",
      T ~ "Nincs"
    ),
```

```r
    words = fct_reorder(words, n)
  ) %>%
  ggplot() +
  aes(n, words, fill = value) +
  geom_vline(xintercept = 0) +
  geom_col(color = "black") +
  labs(title = 'Fordítást követően', x = 'Előfordulási gyakoriság', y = NULL,
       fill = "Adott szó szentimentje") +
  scale_fill_manual(values = c('red4', 'gray70', 'green'))
mygrid <- data.frame(
  row = c(5, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6),
  col = c(7, 1, 3, 4, 7, 7, 5, 4, 2, 3, 7, 2, 3, 5, 4, 4, 7, 6, 2, 5, 3, 6, 4, 5, 2, 4, 7, 1, 6),
  code = c("BG", "IS", "NO", "FI", "EE", "LV", "SE", "DK", "UK", "NL", "LT", "BE", "DE", "PL", "CZ", "A
  name = c("Bulgária", "Izland", "Norvégia", "Finnország", "Észtország", "Lettország", "Svédország", "D
  stringsAsFactors = FALSE
)
geofacet::grid_preview(mygrid)


ggplot(dat_sentiment_daily, aes(date, value)) +
  geom_hline(yintercept = 0, color = "grey20") +
  geom_line(size = 1) +
  facet_geo(~ code, grid = mygrid, label = 'name') +
  scale_x_date(limits = c(min(dat_sentiment_daily$date), max(dat_sentiment_daily$date)),
               breaks = c(min(dat_sentiment_daily$date), max(dat_sentiment_daily$date))) +
  labs(y = "Szentiment", x = NULL)

covid_df <- readr::read_csv("https://covid.ourworldindata.org/data/owid-covid-data.csv")

covid_df %>%
  transmute(name = location, date, cases = new_cases_per_million*1000,
            death = new_deaths_per_million*1000) %>%
  merge(dat) %>%
  select(name, date, cases, death, value) %>%
  pivot_longer(3:5, names_to = "var") %>%
  ggplot(aes(date, value)) +
  geom_line() +
  facet_grid(var ~ name, scales = "free_y")

unemployment <- eurostat::get_eurostat("une_rt_m") %>%
  filter(age == "TOTAL", sex == "T", s_adj == "NSA", unit == "PC_ACT") %>%
  select(geo, time, values) %>%
  mutate(
    year = lubridate::year(time),
    month = lubridate::month(time)
  )

library(reshape2)
dat_sentiment %>%
  na.omit() %>%
  mutate(
    sentiment = ifelse(value > 0, "Pozitív", "Negatív")
  ) %>%
```

```r
  count(words, sentiment, sort = TRUE) %>%
  acast(words ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("red4", "cyan4"),
                   max.words = 100)


dat_sentiment_monthly %>%
  ggplot() +
  geom_hline(yintercept = 0) +
  geom_line(aes(date, value, color = "Szentiment"))  +
  facet_wrap(~code)
```