future_cases

Ngoc Duong

4/30/2020

```
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyverse
```

```
## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------ tidyverse_confli
## x tidyr::complete() masks RCurl::complete()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

```
library(knitr)
```

load curves.RData

```
load("./curves.RData")
```

```
########## Additional days
http = "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_

dates = seq.Date(as.Date("04-21-2020", format = "%m-%d-%Y"),
                 as.Date("04-29-2020", format = "%m-%d-%Y"),
                 by = "day")
dates = format(dates, "%m-%d-%Y")

urls = paste0(http, dates, ".csv")

files = lapply(1:length(dates), function(i){
  d = read.csv(urls[[i]])
  d$date = dates[[i]]

  return(d)
})

covid_update = dplyr::bind_rows(files) %>%
  janitor::clean_names() %>%
  dplyr::select(country_region,
                lat,
                long,
```

```r
                date,
                confirmed,
                deaths) %>%
  rename(confirmed_cases = confirmed,
         fatalities = deaths)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
```

```
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
```

```
## into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector, coercing
## into character vector
```

Some more cleaning

```
covid_update1 = covid_update %>% drop_na() %>%
  mutate(date = lubridate::parse_date_time(x = date,
                                           orders = c("m/d/y", "m-d-Y")))

## group data by country
by_country_update = covid_update1 %>%
  dplyr::group_by(country_region) %>%
  dplyr::mutate(t = as.numeric(difftime(date, min(date), units = "days"))) %>%
  dplyr::group_by(country_region, t) %>%
  dplyr::summarise(confirmed_cases = sum(confirmed_cases), # group by date and take total
            fatalities = sum(fatalities))

by_country_old = by_country %>% group_by(country_region) %>% summarise(max_t = max(t))
param_df1 = rename(param_df1, `country_region` = region)

by_country_updated = merge(by_country_old, by_country_update, by = "country_region")
by_country_updated = merge(by_country_updated, param_df1, by = "country_region") %>%
  mutate(new_t = max_t + t) %>% dplyr::select(-max_t, -t)
```

Find predicted cases using the estimated parameters

```
pred_case = function(a,b,c,t){
  a/(1+exp(-b*(t-c)))
}

by_country_pred = by_country_updated %>%
  mutate(pred_case = pred_case(a,b,c,new_t)) %>%
  group_by(country_region) %>%
  mutate(rmse = sqrt(mean(pred_case - confirmed_cases)^2))


by_country_pred %>% dplyr::select(country_region, rmse) %>%
  rename(`Country` = country_region, RMSE = rmse) %>% distinct(Country, .keep_all = TRUE) %>%
  knitr::kable()
```

Country	RMSE
Afghanistan	5.393774e+02
Albania	2.039204e+02
Algeria	6.387214e+02
Andorra	1.319672e+02
Angola	7.148887e+00
Antigua and Barbuda	1.090799e+01
Argentina	2.700216e+02
Armenia	6.719711e+02
Australia	1.917067e+03
Austria	2.994027e+03
Azerbaijan	7.202949e+02
Bahamas	3.436874e+00

Country	RMSE
Bahrain	1.413542e+02
Bangladesh	6.756127e+01
Barbados	1.240739e+01
Belarus	2.118964e+03
Belgium	6.507217e+03
Belize	1.288055e+01
Benin	2.906674e+00
Bhutan	2.878646e+00
Bolivia	1.725455e+02
Bosnia and Herzegovina	3.949524e+02
Botswana	5.132022e+00
Brazil	1.269550e+04
Brunei	3.629912e+01
Bulgaria	3.208963e+02
Burkina Faso	1.288226e+02
Burma	5.674487e+01
Burundi	3.150043e+00
Cabo Verde	4.907230e+00
Cambodia	2.200000e+01
Cameroon	5.029090e+02
Canada	9.345692e+03
Central African Republic	6.735695e+00
Chad	4.966017e+00
Chile	3.244997e+03
China	3.903014e+03
Colombia	1.212958e+02
Congo (Brazzaville)	2.945546e+01
Congo (Kinshasa)	4.021442e+01
Costa Rica	1.320139e+02
Cote d'Ivoire	1.958214e+02
Croatia	8.517829e+02
Cuba	3.501734e+02
Cyprus	7.838573e+01
Czechia	7.143915e+02
Denmark	3.675242e+01
Djibouti	7.144755e+01
Dominica	4.688070e+00
Dominican Republic	1.408274e+02
Ecuador	9.704089e+03
Egypt	7.779355e+01
El Salvador	3.694334e+00
Equatorial Guinea	6.077510e+01
Eritrea	1.300647e+01
Estonia	6.387457e+02
Eswatini	2.393357e+01
Ethiopia	3.807193e+01
Fiji	5.246838e+00
Finland	5.855149e+02
France	5.996669e+03
Gabon	1.030774e+01
Gambia	3.451035e+00
Georgia	5.344686e+01

Country	RMSE
Germany	9.518427e+03
Ghana	1.978257e+02
Greece	8.678952e+02
Grenada	4.809817e-01
Guatemala	6.746707e+01
Guinea	2.419880e+02
Guinea-Bissau	7.627490e+00
Guyana	4.554371e+00
Haiti	1.044541e+01
Holy See	3.853050e+00
Honduras	1.213332e+02
Hungary	4.918655e+02
Iceland	7.637281e+02
India	4.040879e+03
Indonesia	7.421030e+02
Iran	4.368335e+03
Iraq	4.322925e+02
Ireland	1.314991e+03
Israel	1.245853e+03
Italy	2.035703e+04
Jamaica	3.454265e+01
Japan	4.660723e+02
Jordan	3.503195e+01
Kazakhstan	8.675642e+01
Kenya	1.038083e+01
Korea, South	7.281889e+02
Kosovo	3.100181e+02
Kuwait	1.473255e+02
Kyrgyzstan	4.922672e+01
Laos	6.743800e+00
Latvia	7.756590e+01
Lebanon	1.458798e+02
Liberia	2.495941e+01
Libya	7.722492e+00
Liechtenstein	1.572025e+01
Lithuania	3.973440e+02
Luxembourg	1.466249e+03
Madagascar	2.412224e+01
Malawi	4.213364e+00
Malaysia	1.160317e+03
Maldives	1.198595e+02
Mali	1.385149e+00
Malta	1.174601e+01
Mauritania	4.348619e+00
Mauritius	1.005157e+02
Mexico	2.585209e+03
Moldova	4.890115e+02
Monaco	4.366822e+00
Mongolia	6.659867e+00
Montenegro	8.374034e+01
Morocco	1.622946e+02
Mozambique	6.743138e+00

Country	RMSE
Namibia	6.174278e+00
Nepal	1.113233e-01
Netherlands	1.175214e+03
New Zealand	4.638049e+02
Nicaragua	1.310686e+00
Niger	1.701003e+02
Nigeria	4.819036e+02
North Macedonia	4.678056e+02
Norway	1.072832e+03
Oman	1.976585e+02
Pakistan	3.883162e+03
Panama	2.547662e+02
Papua New Guinea	5.075545e+00
Paraguay	4.631618e+01
Peru	7.100060e+03
Philippines	2.512801e+02
Poland	1.836031e+03
Portugal	3.616519e+03
Qatar	1.383423e+03
Romania	1.474609e+03
Russia	8.262254e+02
Rwanda	1.191240e+01
Saint Kitts and Nevis	4.345856e+00
Saint Lucia	7.288683e+00
Saint Vincent and the Grenadines	6.839551e+00
San Marino	7.778153e+01
Sao Tome and Principe	9.851179e-01
Saudi Arabia	1.084724e+03
Senegal	2.120645e+02
Serbia	1.177727e+03
Seychelles	6.044160e+00
Sierra Leone	1.281400e+01
Singapore	4.830709e+03
Slovakia	2.632267e+02
Slovenia	5.218155e+02
Somalia	2.989741e+02
South Africa	4.022131e+02
South Sudan	5.764170e+00
Spain	2.585860e+04
Sri Lanka	1.375117e+02
Sudan	1.988321e+01
Suriname	5.515508e+00
Sweden	1.873688e+03
Switzerland	4.592913e+02
Syria	8.010362e+00
Taiwan*	3.908048e+01
Tanzania	2.773650e+02
Thailand	2.840538e+03
Timor-Leste	2.152176e+01
Togo	8.101070e-02
Trinidad and Tobago	1.533954e+01
Tunisia	8.565729e+00

Country	RMSE
Turkey	8.599936e+03
Uganda	3.893201e+00
Ukraine	1.071686e+03
United Arab Emirates	1.486016e+03
United Kingdom	1.465031e+04
Uruguay	8.854061e+01
US	1.281198e+05
Uzbekistan	9.055693e+02
Venezuela	3.663289e+01
Vietnam	9.654830e+01
West Bank and Gaza	1.321966e+02
Western Sahara	1.793030e+00
Yemen	8.180930e-02
Zambia	1.493218e+01
Zimbabwe	6.330638e+00

```
by_country_pred %>% ungroup() %>% summarise(`Min RMSE` = min(rmse),
                                            `Max RMSE` = max(rmse),
                                            `Mean RMSE` = mean(rmse),
                                            `Median RMSE` = median(rmse)) %>% knitr::kable()
```

Min RMSE	Max RMSE	Mean RMSE	Median RMSE
0.0810107	128119.8	1867.751	128.8226

Using the estimated parameter for the period until 04/20/2020, we were able to make predictions for the following 8 days until 04/29/2020. The median RMSE for these new predictions across all countries is 131.97. The lowest RMSE is 0.94, and mean RMSE is within acceptable range given the rapid surge in recent number of cases. The maximum RMSE, however, is quite high, at more than 100,000. This is the case of the US. As we did not model each state separately but instead lump all data under the country USA, it is reasonable to see high deviances between the number of predicted cases and actual confirmed cases. As we all know, different states within the US are seeing the number of confirmed cases rising at different rates, on top of discrepancies in testing availability throughout the country, there is no one-size-fits-all model that might capture precisely these underlying trends.

All in all, besides the high RMSE in the US, which is due to vast differences between regions populations and state policies regarding public health interventions, the RMSE for all other countries are within reasonable range, which gives us confidence in the model's ability to predict well future cases.

Export data

```
write.csv(by_country_updated, "./by_country_updated.csv", row.names = FALSE)
```