Practice1 Data Manipulation

Data on COVID-19 (coronavirus) by Our World in Data

description: https://github.com/owid/covid-19-data/blob/master/public/data/README.md

```
owid <- read.csv('https://github.com/owid/covid-19-data/raw/master/public/data/owid-covid-data.csv', stringsAsFactors = F)

dim(owid)

## [1] 105708     60

names(owid)

##  [1] "iso_code"
##  [2] "continent"
##  [3] "location"
##  [4] "date"
##  [5] "total_cases"
##  [6] "new_cases"
##  [7] "new_cases_smoothed"
##  [8] "total_deaths"
##  [9] "new_deaths"
## [10] "new_deaths_smoothed"
## [11] "total_cases_per_million"
## [12] "new_cases_per_million"
## [13] "new_cases_smoothed_per_million"
## [14] "total_deaths_per_million"
## [15] "new_deaths_per_million"
## [16] "new_deaths_smoothed_per_million"
## [17] "reproduction_rate"
## [18] "icu_patients"
## [19] "icu_patients_per_million"
## [20] "hosp_patients"
## [21] "hosp_patients_per_million"
## [22] "weekly_icu_admissions"
## [23] "weekly_icu_admissions_per_million"
## [24] "weekly_hosp_admissions"
## [25] "weekly_hosp_admissions_per_million"
## [26] "new_tests"
## [27] "total_tests"
## [28] "total_tests_per_thousand"
## [29] "new_tests_per_thousand"
## [30] "new_tests_smoothed"
## [31] "new_tests_smoothed_per_thousand"
```

```
## [32] "positive_rate"
## [33] "tests_per_case"
## [34] "tests_units"
## [35] "total_vaccinations"
## [36] "people_vaccinated"
## [37] "people_fully_vaccinated"
## [38] "new_vaccinations"
## [39] "new_vaccinations_smoothed"
## [40] "total_vaccinations_per_hundred"
## [41] "people_vaccinated_per_hundred"
## [42] "people_fully_vaccinated_per_hundred"
## [43] "new_vaccinations_smoothed_per_million"
## [44] "stringency_index"
## [45] "population"
## [46] "population_density"
## [47] "median_age"
## [48] "aged_65_older"
## [49] "aged_70_older"
## [50] "gdp_per_capita"
## [51] "extreme_poverty"
## [52] "cardiovasc_death_rate"
## [53] "diabetes_prevalence"
## [54] "female_smokers"
## [55] "male_smokers"
## [56] "handwashing_facilities"
## [57] "hospital_beds_per_thousand"
## [58] "life_expectancy"
## [59] "human_development_index"
## [60] "excess_mortality"
```

1. column selection

select columns of

iso_code, continent, location, date, new_cases, new_deaths, new_cases_per_million, new_deaths_per_million, reproduction_rate, icu_patients, icu_patients_per_million, hosp_patients, hosp_patients_per_million, weekly_icu_admissions, weekly_icu_admissions_per_million, weekly_hosp_admissions, weekly_hosp_admissions_per_million, new_tests, new_tests_per_thousand, positive_rate, total_vaccinations, people_vaccinated, people_fully_vaccinated, new_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, stringency_index, population, population_density, median_age, aged_65_older, aged_70_older, gdp_per_capita, extreme_poverty, cardiovasc_death_rate, diabetes_prevalence, female_smokers, male_smokers, handwashing_facilities, hospital_beds_per_thousand, life_expectancy, human_development_index, and excess_mortality

from original data.frame of **owid** to make **owid_selected**

```
dim(owid_selected)
```

```
## [1] 105708     43
```

```
names(owid_selected)
```

```
##  [1] "iso_code"                                "continent"
##  [3] "location"                                "date"
##  [5] "new_cases"                               "new_deaths"
##  [7] "new_cases_per_million"                   "new_deaths_per_million"
##  [9] "reproduction_rate"                       "icu_patients"
## [11] "icu_patients_per_million"                "hosp_patients"
## [13] "hosp_patients_per_million"               "weekly_icu_admissions"
## [15] "weekly_icu_admissions_per_million"       "weekly_hosp_admissions"
## [17] "weekly_hosp_admissions_per_million"      "new_tests"
## [19] "new_tests_per_thousand"                  "positive_rate"
## [21] "total_vaccinations"                      "people_vaccinated"
## [23] "people_fully_vaccinated"                 "new_vaccinations"
## [25] "people_vaccinated_per_hundred"           "people_fully_vaccinated_per_hu
ndred"
## [27] "stringency_index"                        "population"
## [29] "population_density"                      "median_age"
## [31] "aged_65_older"                           "aged_70_older"
## [33] "gdp_per_capita"                          "extreme_poverty"
## [35] "cardiovasc_death_rate"                   "diabetes_prevalence"
## [37] "female_smokers"                          "male_smokers"
## [39] "handwashing_facilities"                  "hospital_beds_per_thousand"
## [41] "life_expectancy"                         "human_development_index"
## [43] "excess_mortality"
```

2. excluding non-country location

the location column mostly represents country name. There are some locations for continent-wide or world-wide summary. For simplicity let us exclude non-country locations from **owid_selected** to make **owid_countries**

non-country locations

```
##      iso_code continent       location
## 1    OWID_AFR                  Africa
## 533  OWID_ASI                    Asia
## 1087 OWID_EUR                  Europe
## 1640 OWID_EUN           European Union
## 2193 OWID_INT            International
## 2731 OWID_NAM            North America
## 3285 OWID_OCE                  Oceania
## 3836 OWID_SAM            South America
## 4359 OWID_WRL                    World
```

```
dim(owid_countries)
```

```
## [1] 100796     43
```

```
owid_countries$location %>% unique
```

```
##   [1] "Afghanistan"                    "Albania"

##   [3] "Algeria"                        "Andorra"

##   [5] "Angola"                         "Anguilla"

##   [7] "Antigua and Barbuda"            "Argentina"

##   [9] "Armenia"                        "Aruba"

##  [11] "Australia"                      "Austria"

##  [13] "Azerbaijan"                     "Bahamas"

##  [15] "Bahrain"                        "Bangladesh"

##  [17] "Barbados"                       "Belarus"

##  [19] "Belgium"                        "Belize"

##  [21] "Benin"                          "Bermuda"

##  [23] "Bhutan"                         "Bolivia"
```

```
##  [25] "Bonaire Sint Eustatius and Saba"  "Bosnia and Herzegovina"
##  [27] "Botswana"                         "Brazil"
##  [29] "British Virgin Islands"           "Brunei"
##  [31] "Bulgaria"                         "Burkina Faso"
##  [33] "Burundi"                          "Cambodia"
##  [35] "Cameroon"                         "Canada"
##  [37] "Cape Verde"                       "Cayman Islands"
##  [39] "Central African Republic"         "Chad"
##  [41] "Chile"                            "China"
##  [43] "Colombia"                         "Comoros"
##  [45] "Congo"                            "Cook Islands"
##  [47] "Costa Rica"                       "Cote d'Ivoire"
##  [49] "Croatia"                          "Cuba"
##  [51] "Curacao"                          "Cyprus"
##  [53] "Czechia"                          "Democratic Republic of Congo"
##  [55] "Denmark"                          "Djibouti"
##  [57] "Dominica"                         "Dominican Republic"
##  [59] "Ecuador"                          "Egypt"
##  [61] "El Salvador"                      "Equatorial Guinea"
##  [63] "Eritrea"                          "Estonia"
##  [65] "Eswatini"                         "Ethiopia"
##  [67] "Faeroe Islands"                   "Falkland Islands"
##  [69] "Fiji"                             "Finland"
##  [71] "France"                           "French Polynesia"
```

```
##  [73] "Gabon"                        "Gambia"
##  [75] "Georgia"                      "Germany"
##  [77] "Ghana"                        "Gibraltar"
##  [79] "Greece"                       "Greenland"
##  [81] "Grenada"                      "Guatemala"
##  [83] "Guernsey"                     "Guinea"
##  [85] "Guinea-Bissau"                "Guyana"
##  [87] "Haiti"                        "Honduras"
##  [89] "Hong Kong"                    "Hungary"
##  [91] "Iceland"                      "India"
##  [93] "Indonesia"                    "Iran"
##  [95] "Iraq"                         "Ireland"
##  [97] "Isle of Man"                  "Israel"
##  [99] "Italy"                        "Jamaica"
## [101] "Japan"                        "Jersey"
## [103] "Jordan"                       "Kazakhstan"
## [105] "Kenya"                        "Kiribati"
## [107] "Kosovo"                       "Kuwait"
## [109] "Kyrgyzstan"                   "Laos"
## [111] "Latvia"                       "Lebanon"
## [113] "Lesotho"                      "Liberia"
## [115] "Libya"                        "Liechtenstein"
## [117] "Lithuania"                    "Luxembourg"
```

```
## [119] "Macao"                          "Madagascar"
## [121] "Malawi"                         "Malaysia"
## [123] "Maldives"                       "Mali"
## [125] "Malta"                          "Marshall Islands"
## [127] "Mauritania"                     "Mauritius"
## [129] "Mexico"                         "Micronesia (country)"
## [131] "Moldova"                        "Monaco"
## [133] "Mongolia"                       "Montenegro"
## [135] "Montserrat"                     "Morocco"
## [137] "Mozambique"                     "Myanmar"
## [139] "Namibia"                        "Nauru"
## [141] "Nepal"                          "Netherlands"
## [143] "New Caledonia"                  "New Zealand"
## [145] "Nicaragua"                      "Niger"
## [147] "Nigeria"                        "Niue"
## [149] "North Macedonia"                "Northern Cyprus"
## [151] "Norway"                         "Oman"
## [153] "Pakistan"                       "Palestine"
## [155] "Panama"                         "Papua New Guinea"
## [157] "Paraguay"                       "Peru"
## [159] "Philippines"                    "Pitcairn"
## [161] "Poland"                         "Portugal"
## [163] "Qatar"                          "Romania"
## [165] "Russia"                         "Rwanda"
```

```
## [167] "Saint Helena"                    "Saint Kitts and Nevis"

## [169] "Saint Lucia"                     "Saint Vincent and the Grenadines
"
## [171] "Samoa"                           "San Marino"

## [173] "Sao Tome and Principe"           "Saudi Arabia"

## [175] "Senegal"                         "Serbia"

## [177] "Seychelles"                      "Sierra Leone"

## [179] "Singapore"                       "Sint Maarten (Dutch part)"

## [181] "Slovakia"                        "Slovenia"

## [183] "Solomon Islands"                 "Somalia"

## [185] "South Africa"                    "South Korea"

## [187] "South Sudan"                     "Spain"

## [189] "Sri Lanka"                       "Sudan"

## [191] "Suriname"                        "Sweden"

## [193] "Switzerland"                     "Syria"

## [195] "Taiwan"                          "Tajikistan"

## [197] "Tanzania"                        "Thailand"

## [199] "Timor"                           "Togo"

## [201] "Tonga"                           "Trinidad and Tobago"

## [203] "Tunisia"                         "Turkey"

## [205] "Turkmenistan"                    "Turks and Caicos Islands"

## [207] "Tuvalu"                          "Uganda"

## [209] "Ukraine"                         "United Arab Emirates"

## [211] "United Kingdom"                  "United States"
```

```
## [213] "Uruguay"                          "Uzbekistan"

## [215] "Vanuatu"                          "Vatican"

## [217] "Venezuela"                        "Vietnam"

## [219] "Wallis and Futuna"                "Yemen"

## [221] "Zambia"                           "Zimbabwe"
```

3. changing column names

From column names of new_cases, new_deaths, new_cases_per_million, new_deaths_per_million, new_tests, new_tests_per_thousand, and new_vaccinations,

change the keyword **new** into **daily** to make the meaning more clear.

For example, from **new_cases** to **daily_cases**.

4. change type of column

date column is now in character type. Let us change the type into Date type for further process.

5. proportion of COVID19 contracted people

최근 날짜 기준으로 각 국가별 누적확진자의 전체 인구대비 비율을 계산하여 prop_contr_people column 을 만들자.

예를 들어 인구 1,000,000 인 국가의 누적확진자가 10,000 명이면 확진자 비율은 10000/1000000 = 0.01 => 1% 입니다.

전체 인구 중 가장 높은 percentage 가 감염된 국가는 어느 국가인지 정렬하여 표현하여라. 가장 적은 인구 비율이 감염된 국가는 어디인가?

6. 백신 접종 격차에 대해

뉴스에서는 저개발국가와 선진국 사이의 백신 보급의 양극화에 대해서 우려하는 목소리를 보도하곤 한다.

최근 날짜 기준으로 백신 접종율이 50% 이상인 국가, 10% ~ 50%인 국가, 10% 이하인 국가에 대해서 GDP 의 평균 및 분포를 비교하여보라.

국가 경제 수준에 따른 백신 보급의 편차가 얼마나 큰지 확인하여보라

7. continent comparision

대륙별 총인구수, 평균 인구밀도, GDP, median age 등 여러지표를 비교하여보라.

최근 날짜기준으로 백신접종율 또한 대륙별로 비교하여보라.

무엇을 알 수 있는가?

8. stringency_index v.s. reproduction_rate

stringency_index 는 lockdown 이나 social-distancing 과 같이 감염병 확산을 막기 위한 정부에 의한 사회통제 수준을 의미한다.

0 에서 100 사의 값을 가지며 100 이 가장 강한 수준의 통제를 의미한다.

reproduction_rate 은 한 명의 covid19 확진자가 몇 명의 확진자를 추가로 발생시키는지를 나타내는 지표이다.

높은 수준의 사회적 통제가 전염병의 확산 속도를 줄이는데 긍정적인 역할을 하는지 데이터를 통해서 확인하여보라.