



Data422 Project

Team: GLWY

Introduction

- This project is regarding data sourcing and wrangling that centres on Covid – 19.
- In preparation for data analysis on the impact of Covid-19 on economy, education and demographic aspects.

Data Source

- WHO Coronavirus Disease Dataset:
- <https://covid19.who.int/>
- World Bank Open Data:
- <https://data.worldbank.org/>
- OCHA Services Dataset: HUMANITARIAN DATA EXCHANGE
- <https://data.humdata.org/dataset>

Process & Delivered Dataset

```
#Check country code and Unified country code
data_WDICountry_code<-data_WDICountry_raw %>%
  rename(Country_Code = "Country Code",Country_Code_2="2-alpha code", Table_Name="Table Name")
data_WDICountry_code<-data_WDICountry_code %>%
  select(Country_Code,Country_Code_2,Table_Name)
data_GEPData <- data_GEPData_raw %>% rename(Country_Code_gep = "Country Code",Country_Name = "Country Name")|
  data_GEPData<-data_GEPData %>% select(Country_Name,Country_Code_gep)
```

#Country code in GEP data set not match in WDICountry data set

```
result <-full_join(data_WDICountry_code,data_GEPData,by=c("Country_Code"="Country_Code_gep"))
result %>%
  filter(is.na(Table_Name))
```

Country_Code	Country_Code_2	Table_Name	Country_Name
<chr>	<chr>	<chr>	<chr>
AME	NA	NA	Advanced economies
EAA	NA	NA	East Asia and Pacific
EMD	NA	NA	Emerging Markets and Developing Economies (EMDE)
E19	NA	NA	Euro Area
ECH	NA	NA	Europe and Central Asia
LAP	NA	NA	Latin America and the Caribbean
MNH	NA	NA	Middle East and North Africa
SAP	NA	NA	South Asia

- Due to the different sources of the data, we need to match the country codes.
- We select the country code of WDI Country dataset as the basis.

Join all country code data in one dataframe
 Find out the different country codes
 Use function to replace them with the same country codes.

Country_Code	Country_Name.x	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
EAS	East Asia & Pacific	7.0698925	4.6104777	4.6941374	4.7698198	4.1898646	4.19974920	4.0914041	6.5	6.3	5.9	0.5
EMU	Euro area	2.1335091	1.6865703	-0.8921201	-0.2527230	1.4010985	2.11356295	1.9128129	2.5	1.9	1.2	-9.1
ECS	Europe & Central Asia	2.6224686	2.4046841	0.3183861	0.9001904	1.8326425	2.07578414	1.9211933	4.1	3.3	2.2	-4.7
LCN	Latin America & Caribbean	5.8501657	4.3658387	2.7841570	2.7848736	0.9906779	0.09008685	-0.3367020	1.9	1.7	0.8	-7.2
MEA	Middle East & North Africa	5.0516374	3.6810002	3.8893955	2.7076935	2.9054469	2.37541872	4.9624285	1.1	0.9	-0.2	-4.2
SAS	South Asia	7.7035003	5.1375769	5.5019859	6.0875529	6.9926022	7.48032879	7.7783585	6.5	6.5	4.7	-2.7
SSF	Sub-Saharan Africa	5.4048842	4.7380671	3.9640744	5.0013742	4.6562645	2.84441682	1.2371439	2.6	2.6	2.2	-2.8

```
#Convert country code
convert <- function(country){
  if (country %in% "EAA"){
    country = "EAS"
  }
  if (country %in% "E19"){
    country = "EMU"
  }
  if (country %in% "ECH"){
    country = "ECS"
  }
  if (country %in% "LAP"){
    country = "LCN"
  }
  if (country %in% "MNH"){
    country = "MEA"
  }
  if (country %in% "SAP"){
    country = "SAS"
  }
  if (country %in% "SSP"){
    country = "SSF"
  }
  if (country %in% "WLT"){
    country = "WLD"
  }
  else country = country
}
```

```

which(mydata_realGDP[,1]=="World") # No.4 row
which(mydata_realGDP[,1]=="Non-energy commodity price index") # No.47 row

mydata_realGDP<-mydata_realGDP[3:47,]

```

```

mydata_realGDP<-mydata_realGDP[3:47,]
#Check NA Columns
which(is.na(mydata_realGDP[1,])) # No.2-5,11 columns
mydata_realGDP1 <-mydata_realGDP[-c(2:5,11)]
mydata_realGDP1<-rename(mydata_realGDP1,classification="TABLE 1.1 Real GDP1","2017"="...6","2018"="...7","2019e"="...8","2020f"=
mydata_realGDP1_1<-mydata_realGDP1[2:45,]
mydata_realGDP1_1

```

- Import data into R basing on each Tab
- According to the key words, use the which function to get data position.
- Combine all tabs data
- Remove NA and Rename columns' name

classification	2017	2018	2019e	2020f	2021f	2020fd	2021fd
	<chr>	<dbl>		<chr>	<chr>	<chr>	<chr>
World	3.3	3.0	2.4	-5.2	4.2	-7.7	1.6
Advanced economies	2.5	2.1	1.6	-7	3.9	-8.4	2.4
United States	2.4	2.9	2.2999999999999998	-6.1	4	-7.9	2.2999999999999998
Euro Area	2.5	1.9	1.2	-9.1	4.5	-10.1	3.2
Japan	2.2	0.3	0.7	-6.1	2.5	-6.8	1.9
Emerging market and developing economies	4.5	4.3	3.5	-2.5	4.599999999999996	-6.6	0.3
Commodity-exporting EMDEs	2.2	2.1	1.5	-4.8	3.1	-7.4	0.2
Other EMDEs	6.1	5.7	4.8	-1.100000000000001	5.5	-6.2	0.3
Other EMDEs excluding China	5.4	4.8	3.2	-3.6	3.6	-7.6	-0.8
East Asia and Pacific	6.5	6.3	5.9	0.5	6.6	-5.2	1

- Some of the data contains subscripts and comma
- Use regular expressions to remove subscripts and comma

```
#Tidy classification information
```

```
mydata_RealGDP$classification<-sub("[0-9.*]",replacement="",sub("[0-9]","",sub("[\\,]","",mydata_RealGDP$classification,perl=T)))
mydata_RealGDP
```

Egypt2	4.2	5.3	5.6	3	2.1	-2.8	-3.9
South Asia	6.5	6.5	4.7	-2.7	2.8	-8.199999999999993	-3.1
India3	7.0	6.1	4.2	-3.2	3.1	-9	-3
Pakistan2	5.2	5.5	1.9	-2.6	-0.2	-5	-3.2
Bangladesh2	7.3	7.9	8.199999999999993	1.6	1	-5.6	-6.3
Sub-Saharan Africa	2.6	2.6	2.2000000000000002	-2.8	3.1	-5.8	0
Nigeria	0.8	1.9	2.2000000000000002	-3.2	1.7	-5.3	-0.4
South Africa	1.4	0.8	0.2	-7.1	2.9	-8	1.6

Reshaping data frame from wide to long

```
GDP_Area <-data_GDP_His %>%
filter(Country_Code=="EAS" | Country_Code=="ECS" | Country_Code=="LCN" | Country_Code=="MEA" | Country_Code=="SAS" | Country_Code=="SSF")
data_long <- gather(GDP_Area, year, GDP, "2010":"2021", factor_key=TRUE)
data_long$year<- as.Date(data_long$year,"%Y")
str(data_long)
```

A tibble: 6 × 14													
Country_Code	Country_Name.x	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
EAS	East Asia & Pacific	7.069893	4.610478	4.6941374	4.7698198	4.1898646	4.19974920	4.091404	6.5	6.3	5.9	0.5	6.6
ECS	Europe & Central Asia	2.622469	2.404684	0.3183861	0.9001904	1.8326425	2.07578414	1.921193	4.1	3.3	2.2	-4.7	3.6
LCN	Latin America & Caribbean	5.850166	4.365839	2.7841570	2.7848736	0.9906779	0.09008685	-0.336702	1.9	1.7	0.8	-7.2	2.8
MEA	Middle East & North Africa	5.051637	3.681000	3.8893955	2.7076935	2.9054469	2.37541872	4.962428	1.1	0.9	-0.2	-4.2	2.3
SAS	South Asia	7.703500	5.137577	5.5019859	6.0875529	6.9926022	7.48032879	7.778358	6.5	6.5	4.7	-2.7	2.8
SSF	Sub-Saharan Africa	5.404884	4.738067	3.9640744	5.0013742	4.6562645	2.84441682	1.237144	2.6	2.6	2.2	-2.8	3.1

A tibble: 72 × 4			
Country_Code	Country_Name.x	year	GDP
<chr>	<chr>	<date>	<dbl>
EAS	East Asia & Pacific	2010-10-13	7.0698925
ECS	Europe & Central Asia	2010-10-13	2.6224686
LCN	Latin America & Caribbean	2010-10-13	5.8501657
MEA	Middle East & North Africa	2010-10-13	5.0516374
SAS	South Asia	2010-10-13	7.7035003
SSF	Sub-Saharan Africa	2010-10-13	5.4048842
EAS	East Asia & Pacific	2011-10-13	4.6104777
ECS	Europe & Central Asia	2011-10-13	2.4046841
LCN	Latin America & Caribbean	2011-10-13	4.3658387
MEA	Middle East & North Africa	2011-10-13	3.6810002
SAS	South Asia	2011-10-13	5.1375769
SSF	Sub-Saharan Africa	2011-10-13	4.7380671
EAS	East Asia & Pacific	2012-10-13	4.6941374
ECS	Europe & Central Asia	2012-10-13	0.3183861
LCN	Latin America & Caribbean	2012-10-13	0.7044570

GDP history from 2010 to 2021

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Country_C	Country_N	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
EAS	East Asia & Pacific	7.069893	4.610478	4.694137	4.76982	4.189865	4.199749	4.091404	6.5	6.3	5.9	0.5	6.6
EMU	Euro area	2.133509	1.68657	-0.89212	-0.25272	1.401099	2.113563	1.912813	2.5	1.9	1.2	-9.1	4.5
ECS	Europe & Central Asia	2.622469	2.404684	0.318386	0.90019	1.832642	2.075784	1.921193	4.1	3.3	2.2	-4.7	3.6
LCN	Latin America & the Caribbean	5.850166	4.365839	2.784157	2.784874	0.990678	0.090087	-0.3367	1.9	1.7	0.8	-7.2	2.8
MEA	Middle East & North Africa	5.051637	3.681	3.889396	2.707693	2.905447	2.375419	4.962428	1.1	0.9	-0.2	-4.2	2.3
SAS	South Asia	7.7035	5.137577	5.501986	6.087553	6.992602	7.480329	7.778358	6.5	6.5	4.7	-2.7	2.8
SSF	Sub-Saharan Africa	5.404884	4.738067	3.964074	5.001374	4.656265	2.844417	1.237144	2.6	2.6	2.2	-2.8	3.1
WLD	World	4.300652	3.140277	2.516963	2.662936	2.847459	2.879152	2.592172	3.3	3	2.4	-5.2	4.2
AEG	Afghanistan	14.36244	0.126255	12.75220	5.600745	2.721512	1.151315	2.260214	2.7	1.8	2.0	-5.5	1

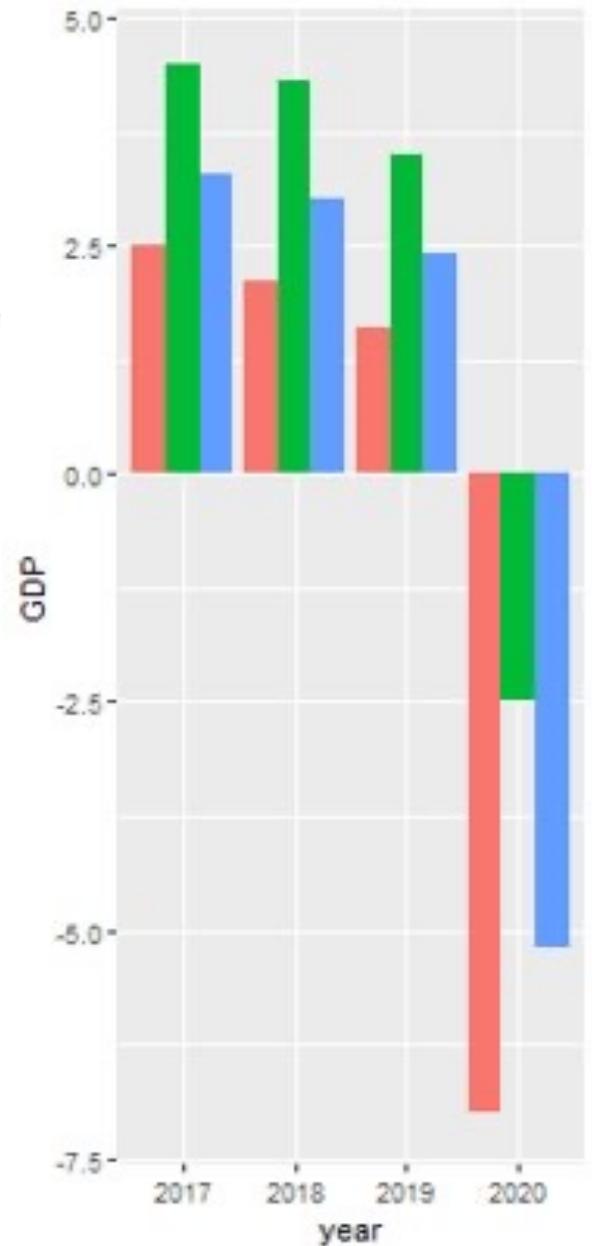
Estimate GDP and Forecast GDP by Area

A	B	C	D	E	F	G	H	I
classification	2017	2018	2019e	2020f	2021f	2020fd	2021fd	area
Cambodia	7	7.5	7.1	-1	6	-7.8	-0.8	EAS
China	6.8	6.6	6.1	1	6.9	-4.9	1.1	EAS
Fiji	5.4	3.5	1	-4.3	1.9	-6	-1	EAS
Indonesia	5.1	5.2	5	0	4.8	-5.1	-0.4	EAS
Lao PDR	6.9	6.3	4.7	1	4.6	-4.8	-1.1	EAS
Malaysia	5.7	4.7	4.3	-3.1	6.9	-7.6	2.4	EAS
Mongolia	5.3	6.9	4.8	-0.5	4.9	-6	-0.3	EAS
Timor-Leste	6.0	6.0	6.0	4.5	6	5.0	0.0	EAS

Summary of Economic Indicators

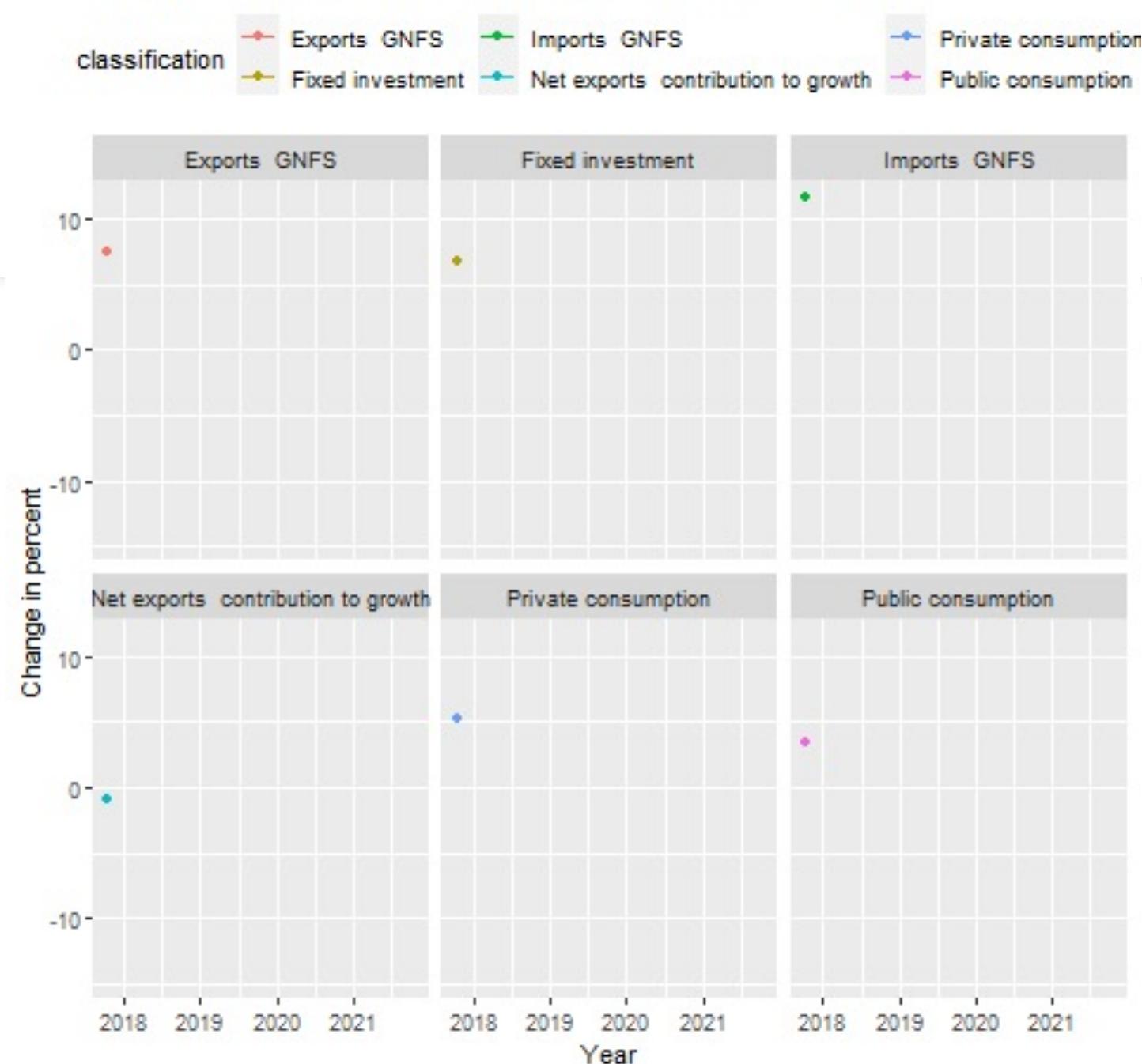
A	B	C	D	E	F	G	H	I
classification	2017	2018	2019e	2020f	2021f	2020fd	2021fd	area
EMDE EAP GDP	6.5	6.3	5.9	0.5	6.6	-5.2	1	EAS
GDP per capita (US. dollars)	5.8	5.6	5.2	-0.1	6	-5.2	1	EAS
EMDE EAP GDP	6.5	6.3	5.9	0.5	6.6	-5.2	1	EAS
PPP GDP	6.4	6.3	5.8	0.4	6.5	-5.3	0.9	EAS
Private consumption	6.1	8.4	6.5	0.8	8.8	-6.1	2.2	EAS
Public consumption	8.9	8.8	7.8	11.2	7.4	3.6	-0.1	EAS
Fixed investment	4.7	5.1	4.3	-0.2	4.1	-4.8	-0.6	EAS
Exports GNFS	9.4	4.9	1.9	-10.3	4.2	-11.6	2.2	EAS
Imports GNFS	8.3	8.4	0.3	-5.7	5.2	-7.7	2.7	EAS
Net exports contribution to	0.4	-0.9	0.5	-1.3	-0.3	-1.1	-0.2	EAS
East Asia excluding China	5.4	5.3	4.8	-1.2	5.4	-6.1	0.4	EAS

Growth in World, advanced economies and EMDEs



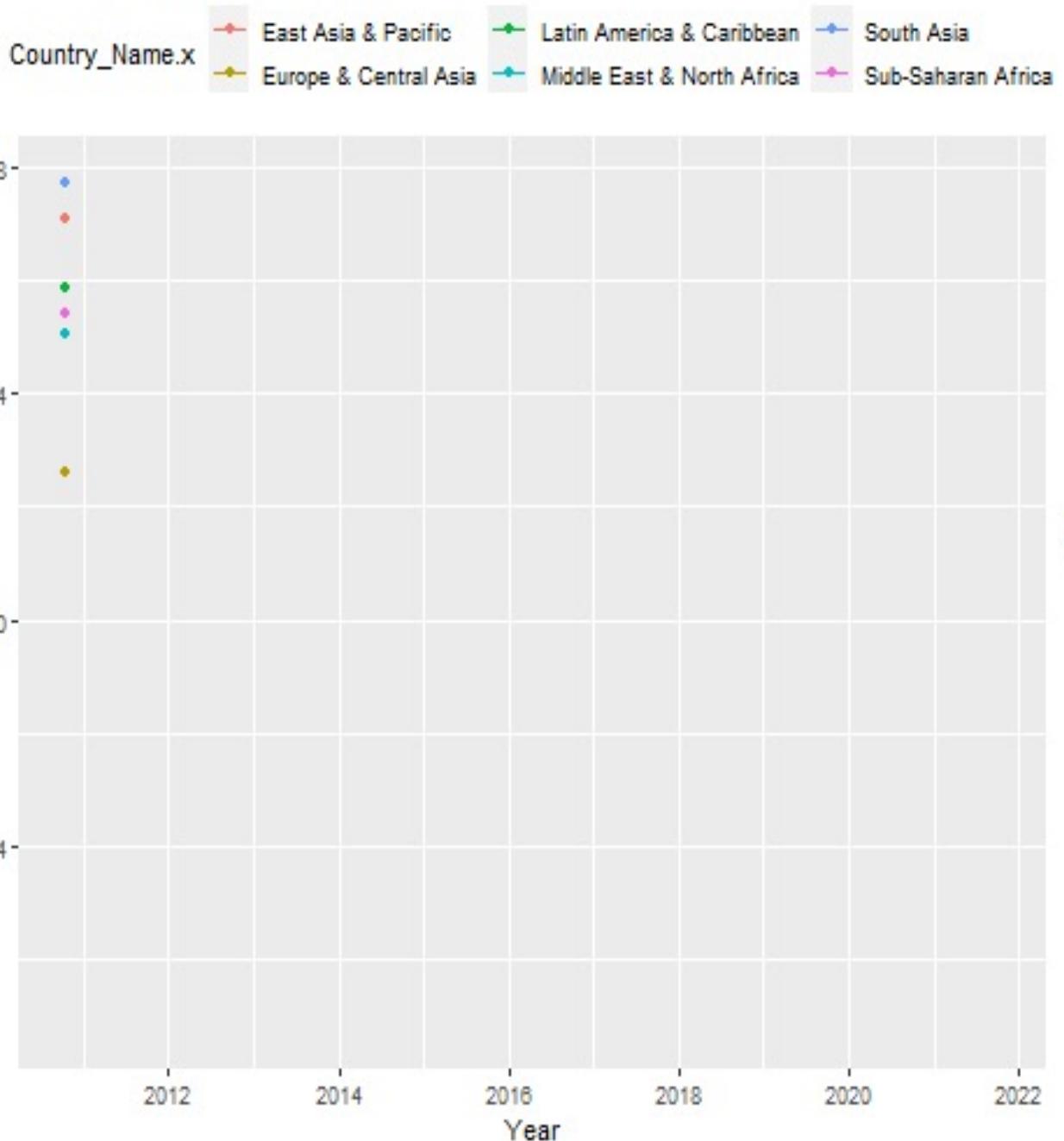
GDP growth is negative in 2020,
especially in developed economies

Europe and Central Asia forecast summary from 2017 to 2021

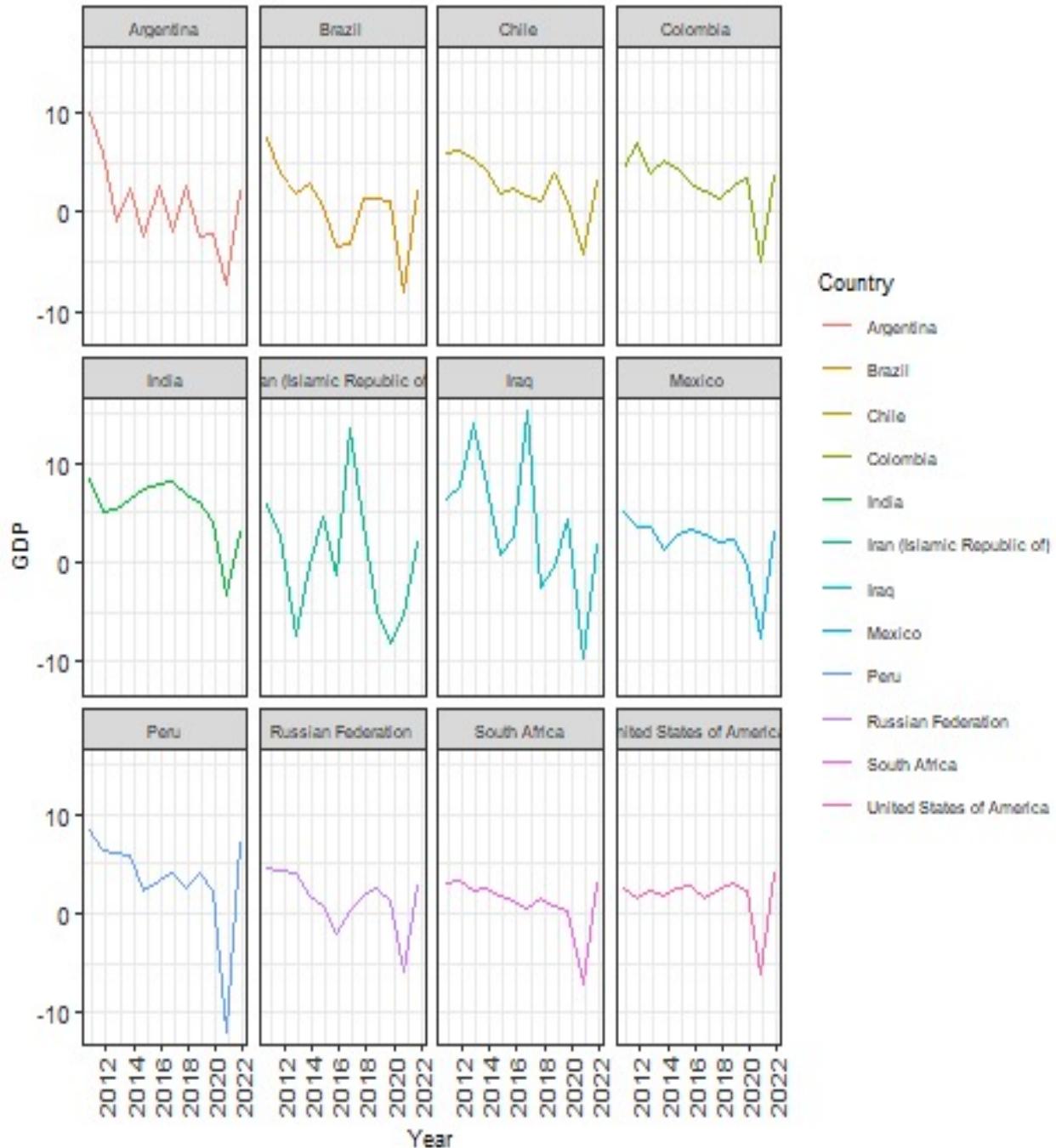


Four of the indicators dropped significantly in 2020

GDP by Area from 2010 to 2021



GDP of the country with the most confirmed Covid-19 cases from 2010-2021



Covid-19 dataframe

- Translate WHO region

```
# translate the WHO regions name
covid_df <- covid_df %>%
  mutate(WHO_region = ifelse(WHO_region == "AMRO", "Americas",
    ifelse(WHO_region == "SEARO", "South-East Asia",
      ifelse(WHO_region == "EURO", "European",
        ifelse(WHO_region == "AFRO", "African",
          ifelse(WHO_region == "WPRO", "Western Pacific",
            ifelse(WHO_region == "EMRO", "Eastern Mediterranean",
              WHO_region
            )
          )
        )
      )
    )
  )
)
```

- Reshape to long

```
# reshape the covid dataframe to a long table with all the numeric values in one column
covid_long <- covid_df %>% pivot_longer(cols=c(New_cases, Cumulative_cases, New_deaths, Cumulative_deaths),
  names_to = "Stat",
  values_to = "Value")
```

A tibble: 260872 × 6

Date	Country_code	Country	WHO_region	Stat	Value
<date>	<chr>	<chr>	<chr>	<chr>	<dbl>
2020-01-03	AF	Afghanistan	Eastern Mediterranean	New_cases	0
2020-01-03	AF	Afghanistan	Eastern Mediterranean	Cumulative_cases	0
2020-01-03	AF	Afghanistan	Eastern Mediterranean	New_deaths	0
2020-01-03	AF	Afghanistan	Eastern Mediterranean	Cumulative_deaths	0

Population dataframe

- Deal with other names in country column

```
'Africa' · 'African Group' · 'African Union' · 'African Union: Central Africa' · 'African Union: Eastern Africa' · 'African Union: Northern Africa' ·
'African Union: Southern Africa' · 'African Union: Western Africa' · 'African, Caribbean and Pacific (ACP) Group of States' · 'Andean Community' ·
'Asia' · 'Asia-Pacific Economic Cooperation (APEC)' · 'Asia-Pacific Group' · 'Association of Southeast Asian Nations (ASEAN)' ·
'Australia/New Zealand' · 'BRIC' · 'BRICS' · 'Belt-Road Initiative (BRI)' · 'Belt-Road Initiative: Africa' · 'Belt-Road Initiative: Asia' ·
'Belt-Road Initiative: Europe' · 'Belt-Road Initiative: Latin America and the Caribbean' · 'Belt-Road Initiative: Pacific' ·
'Black Sea Economic Cooperation (BSEC)' · 'Bolivarian Alliance for the Americas (ALBA)' · 'Caribbean' ·
'Caribbean Community and Common Market (CARICOM)' · 'Central America' · 'Central Asia' · 'Central European Free Trade Agreement (CEFTA)' ·
'Central and Southern Asia' · 'Channel Islands' · 'China (and dependencies)' · 'China, Hong Kong SAR' · 'China, Macao SAR' ·
'China, Taiwan Province of China' · 'Commonwealth of Independent States (CIS)' · 'Commonwealth of Nations' · 'Commonwealth: Africa' ·
'Commonwealth: Asia' · 'Commonwealth: Caribbean and Americas' · 'Commonwealth: Europe' · 'Commonwealth: Pacific' ·
'Countries with Access to the Sea' · 'Countries with Access to the Sea: Africa' · 'Countries with Access to the Sea: Asia' ·
'Countries with Access to the Sea: Europe' · 'Countries with Access to the Sea: Latin America and the Caribbean' ·
'Countries with Access to the Sea: Northern America' · 'Countries with Access to the Sea: Oceania' · 'Côte d'Ivoire' ·
'Dem. People's Republic of Korea' · 'Denmark (and dependencies)' · 'ECE: North America-2' · 'ECE: UNECE-52' · 'ECLAC: Latin America' ·
'ECLAC: The Caribbean' · 'ESCAP region: East and North-East Asia' · 'ESCAP region: North and Central Asia' · 'ESCAP region: Pacific' ·
'ESCAP region: South and South-West Asia' · 'ESCAP region: South-East Asia' · 'ESCAP: ADB Developing member countries (DMCs)' ·
```

```
# creat a list with irrelaveant country names
area_list = list("Europe", "Asia", "WHO", "South America", "UNICEF", "WB", "UNFPA", "Oceania", "Group", "Country",
  "Union", "Organization", "Caribbean", "Pacific", "Low", "Land", "Countries", "SIDS", "ESCAP",
  "ESCWA", "Community", "Latin", "LLDC", "/", "developed", "countries", "League", "BRI", "Western",
  "World", "Member", "dependencies", "Eastern", "Organisation", "group", "Common", "Trade", "ECE")
```

```
# filter out the irrelaveant rows by match the area list
population_df <- population_df %>%
  filter(!grepl(paste(area_list, collapse = "|"), Country))
```

Join Covid-19 and Population dataframe

- One country in each row

```
else if (country %in% "China, Hong Kong SAR"){
  country = "China"
}
else if (country %in% "China, Macao SAR"){
  country = "China"
}
else if (country %in% "China, Taiwan Province of China"){
  country = "China"
}
```

LocID	Country	VarID	Variant	Time	MidPeriod	PopMale	PopFemale	PopTotal	PopDensity
<dbl>	<chr>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
156	China	2	Medium	2020	2020.5	738247.340	701076.434	1439323.774	153.312
344	China	2	Medium	2020	2020.5	3439.666	4057.322	7496.988	7139.989
446	China	2	Medium	2020	2020.5	312.100	337.242	649.342	21717.124
158	China	2	Medium	2020	2020.5	11834.313	11982.462	23816.775	672.600

```
# since multiple rows with same country names, we group them together with one country one row with corresponding
# figures . We can't group population desity with simple calculation, we need to calculate corresponding square km
# first, then calculate density back by the sum of population and sum of square km, also fixing the comma messed up
# with decimal problem when we load the dataset into dataframe
population_df <- population_df %>%
  mutate(PopTotal = PopTotal * 1000) %>%
  mutate(PopDensity = PopDensity * 1000) %>%
  mutate(square_km = PopTotal / PopDensity) %>%
  group_by(Country) %>%
  summarise(Pop_total = sum(PopTotal), PopDensity = Pop_total / sum(square_km))
```

Country	Pop_total	PopDensity
<chr>	<dbl>	<dbl>
China	1471286879	156109.8

Julia with Unemployment Rate dataset

- [Extract unemployment data from HNP dataset](#)

```
# select the useful rows and columns we need
ue_df = ue_df[ue_df[:,3] .== "Unemployment, total (% of total labor force)",
 [1, 2, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65]]
```

```
# rename the columns with more suitable names
rename!(covid_df, Symbol("Date_reported")=>Symbol("Date"), Symbol("Country")=>Symbol("Country"),
 Symbol("WHO_region")=>Symbol("WHO_region"))
```

259 rows × 12 columns (omitted printing of 8 columns)

	Country Name	Country Code	2011	2012
	String	String	Float64?	Float64?
1	Arab World	ARB	10.2847	10.4908
2	Caribbean small states	CSS	10.4141	11.0101
3	Central Europe and the Baltics	CEB	9.78936	9.97771
4	Early-demographic dividend	EAR	5.66664	5.64458
5	East Asia & Pacific	EAS	4.1598	4.07486
6	East Asia & Pacific (excluding high income)	EAP	4.13901	4.06279

- [Reshape to long](#)

```
# reshape the dataframe to a long dataframe
unemployment_long = stack(unemployment_df, ["20$i" for i in 11:20])
```

- [Extract WHO region from Covid-19 dataset](#)

```
# we only need the WHO regions information with corresponding countries from this table
area_df = by(DataFrame(covid_df[:,[3, 4]]), :Country) do group
    DataFrame(WHO_region = group[:, :WHO_region][1])
end
```

235 rows × 2 columns

```
# create a function to translate WHO region name
function translate_region(region)
    if region == "AMRO"
        region = "Americas"
    elseif region == "SEARO"
        region = "South-East Asia"
    elseif region == "EURO"
        region = "European"
    elseif region == "AFRO"
        region = "African"
    elseif region == "WPRO"
        region = "Western Pacific"
    elseif region == "EMRO"
        region = "Eastern Mediterranean"
    else
        region = region
    end
end
```

	Country	WHO_region
	String	String
1	Afghanistan	EMRO
2	Albania	EURO
3	Algeria	AFRO
4	American Samoa	WPRO
5	Andorra	EURO
6	Angola	AFRO

Join unemployment with WHO region data frame by country

- Manually correct unmatched countries

["Bahamas, The", "Bolivia", "Channel Islands", "Congo, Dem. Rep.", "Congo, Rep.", "Côte d'Ivoire", "Czech Republic", "Egypt, Arab Rep.", "Gambia, The", "Hong Kong SAR, China", "Iran, Islamic Rep.", "Korea, Dem. People's Rep.", "Korea, Rep.", "Kyrgyz Republic", "Lao PDR", "Macao SAR, China", "Moldova", "Slovak Republic", "St. Lucia", "St. Vincent and the Grenadines", "Tanzania", "United Kingdom", "United States", "Venezuela, RB", "Vietnam", "Virgin Islands (U.S.)", "West Bank and Gaza", "Yemen, Rep."]

["American Samoa", "Andorra", "Anguilla", "Antigua and Barbuda", "Aruba", "Bahamas", "Bermuda", "Bolivia (Plurinational State of)", "Bonaire, Sint Eustatius and Saba", "British Virgin Islands", "Cayman Islands", "Congo", "Cook Islands", "Côte d'Ivoire", "Curaçao", "Czechia", "Democratic People's Republic of Korea", "Democratic Republic of the Congo", "Dominica", "Egypt", "Falkland Islands (Malvinas)", "Faroe Islands", "French Guiana", "Gambia", "Gibraltar", "Greenland", "Grenada", "Guadeloupe", "Guernsey", "Holy See", "Iran (Islamic Republic of)", "Isle of Man", "Jersey", "Kiribati", "Kosovo[1]", "Kyrgyzstan", "Lao People's Democratic Republic", "Liechtenstein", "Marshall Islands", "Martinique", "Mayotte", "Micronesia (Federated States of)", "Monaco", "Montserrat", "Nauru", "Niue", "Northern Mariana Islands (Commonwealth of the)", "occupied Palestinian territory, including east Jerusalem", "Other", "Palau", "Pitcairn Islands", "Republic of Korea", "Republic of Moldova", "Réunion", "Saint Barthélemy", "Saint Helena", "Saint Kitts and Nevis", "Saint Lucia", "Saint Martin", "Saint Pierre and Miquelon", "Saint Vincent and the Grenadines", "San Marino", "Seychelles", "Sint Maarten", "Slovakia", "The United Kingdom", "Tokelau", "Turks and Caicos Islands", "Tuvalu", "United Republic of Tanzania", "United States of America", "United States Virgin Islands", "Venezuela (Bolivarian Republic of)", "Viet Nam", "Wallis and Futuna", "Yemen"]

- One country code with one set of unemployment figures each row

Country	ISO	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	WHO_region
"China"	"CHN"	4.548	4.57	4.59	4.609	4.629	4.535	4.441	4.276	4.32	4.37	"Western Pacific"
"China"	"HKG"	3.42	3.29	3.4	3.3	3.32	3.39	3.09	2.903	3.629	4.265	"Western Pacific"
"China"	"MAC"	2.56	1.97	1.82	1.7	1.81	1.9	2.0	1.966	2.237	2.505	"Western Pacific"

```
# create a function to amend the country names to get join again
function correct_country(name)
  if name == "Bahamas, The"
    name = "Bahamas"
  elseif name == "Bolivia"
    name = "Bolivia (Plurinational State of)"
  elseif name == "Congo, Dem. Rep."
    name = "Congo"
  elseif name == "Congo, Rep."
    name = "Democratic Republic of the Congo"
  elseif name == "Côte d'Ivoire"
    name = "Côte d'Ivoire"
  elseif name == "Czech Republic"
    name = "Czechia"
  elseif name == "Egypt, Arab Rep."
    name = "Egypt"
  elseif name == "Gambia, The"
    name = "Gambia"
```

```
# create a function to change their country names back after the join is done
function correct_back(ISO, country)
  if ISO == "HKG"
    country = "China HongKong"
  elseif ISO == "MAC"
    country = "China Macao"
  else
    country = country
  end
```

Deliverable Dataset

- Covid-19 with population and education status

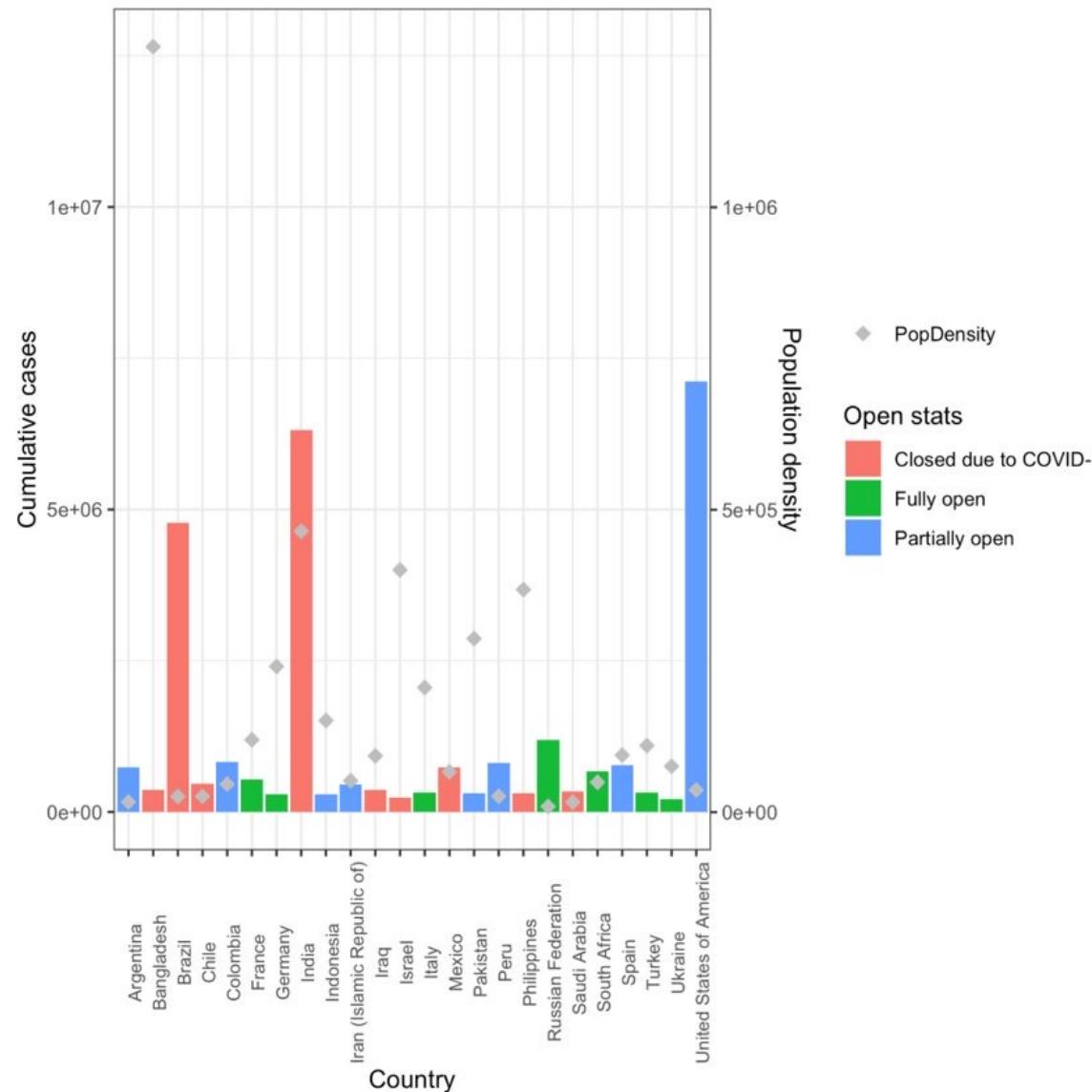
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Date	Country_code	Country	WHO_region	New_cases	Cumulative_case	New_deaths	Cumulative_deaths	Pop_total	PopDensity	CumCasePop_percent	ISO	Status		
82	81	2020/3/23	AF	Afghanistan	Eastern Mediterranean	6	40	1	1	38928341	59627	0.000102753	AFG	Closed due to COVID-19	
83	82	2020/3/24	AF	Afghanistan	Eastern Mediterranean	2	42	0	1	38928341	59627	0.000107891	AFG	Closed due to COVID-19	
84	83	2020/3/25	AF	Afghanistan	Eastern Mediterranean	32	74	0	1	38928341	59627	0.000190093	AFG	Closed due to COVID-19	
85	84	2020/3/26	AF	Afghanistan	Eastern Mediterranean	6	80	1	2	38928341	59627	0.000205506	AFG	Closed due to COVID-19	
86	85	2020/3/27	AF	Afghanistan	Eastern Mediterranean	11	91	0	2	38928341	59627	0.000233763	AFG	Closed due to COVID-19	
87	86	2020/3/28	AF	Afghanistan	Eastern Mediterranean	15	106	1	3	38928341	59627	0.000272295	AFG	Closed due to COVID-19	
88	87	2020/3/29	AF	Afghanistan	Eastern Mediterranean	8	114	1	4	38928341	59627	0.000292846	AFG	Closed due to COVID-19	
89	88	2020/3/30	AF	Afghanistan	Eastern Mediterranean	0	114	0	4	38928341	59627	0.000292846	AFG	Closed due to COVID-19	
90	89	2020/3/31	AF	Afghanistan	Eastern Mediterranean	52	166	0	4	38928341	59627	0.000426425	AFG	Closed due to COVID-19	
91	90	2020/4/1	AF	Afghanistan	Eastern Mediterranean	26	192	0	4	38928341	59627	0.000493214	AFG	Closed due to COVID-19	
92	91	2020/4/2	AF	Afghanistan	Eastern Mediterranean	43	235	0	4	38928341	59627	0.000603673	AFG	Closed due to COVID-19	

- Unemployment

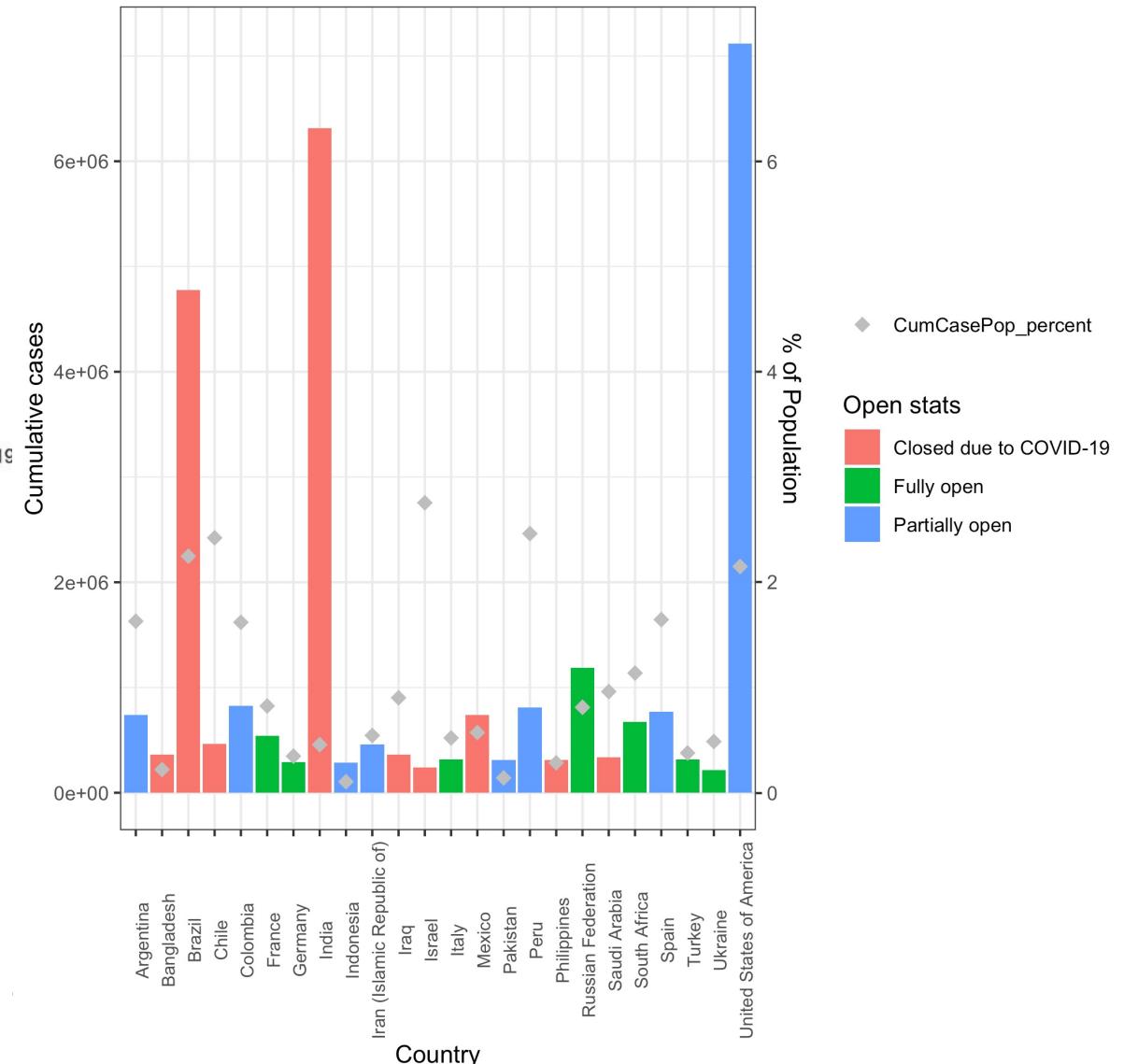
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Country	ISO	WHO_region	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	19-20Diff
2	Afghanistan	AFG	Eastern Mediterranean	11.4879999	11.5080004	11.5340004	11.448	11.3870001	11.3129997	11.184	11.0570002	11.118	11.1639996	0.04599953
3	Albania	ALB	European	13.4809999	13.3760004	15.8660002	17.4899998	17.0799999	15.2200003	13.75	12.3400002	12.3310003	12.8129997	0.4819994
4	Algeria	DZA	African	9.96000004	10.9700003	9.81999969	10.2069998	11.2060003	10.2019997	11.9960003	11.882	11.7040005	11.5249996	-0.1790009
5	Angola	AGO	African	7.36199999	7.37900019	7.4000001	7.33099985	7.28200006	7.22300005	7.11899996	7.01900005	6.88600016	6.77400017	-0.112
6	Argentina	ARG	Americas	7.17999983	7.21700001	7.0999999	7.26800013	7.75	7.97700024	8.34700012	9.22000027	9.78899956	10.4359999	0.64700031
7	Armenia	ARM	European	18.4430008	17.3010006	16.1790009	17.4979992	18.2609997	17.6170006	17.7049999	17.4960003	16.9899998	16.625	-0.3649998
8	Australia	AUS	Western Pacific	5.08099985	5.22399998	5.66200018	6.07700014	6.0539999	5.71000004	5.59100008	5.29799986	5.26800013	5.32800007	0.05999994
9	Austria	AUT	European	4.56400013	4.86499977	5.33500004	5.61999989	5.72300005	6.01399994	5.49900007	4.84800005	4.67399979	4.77899981	0.10500002
10	Azerbaijan	AZE	European	5.42000008	5.19000006	4.96999979	4.90999985	4.96000004	5	5	4.9000001	5.50899982	5.96000004	0.45100021

- The impact of cumulative cases, population density and total cases in percentage of population on status of educational organisations in October

Education Status vs Pop Density in Most Severe Countries in Oct

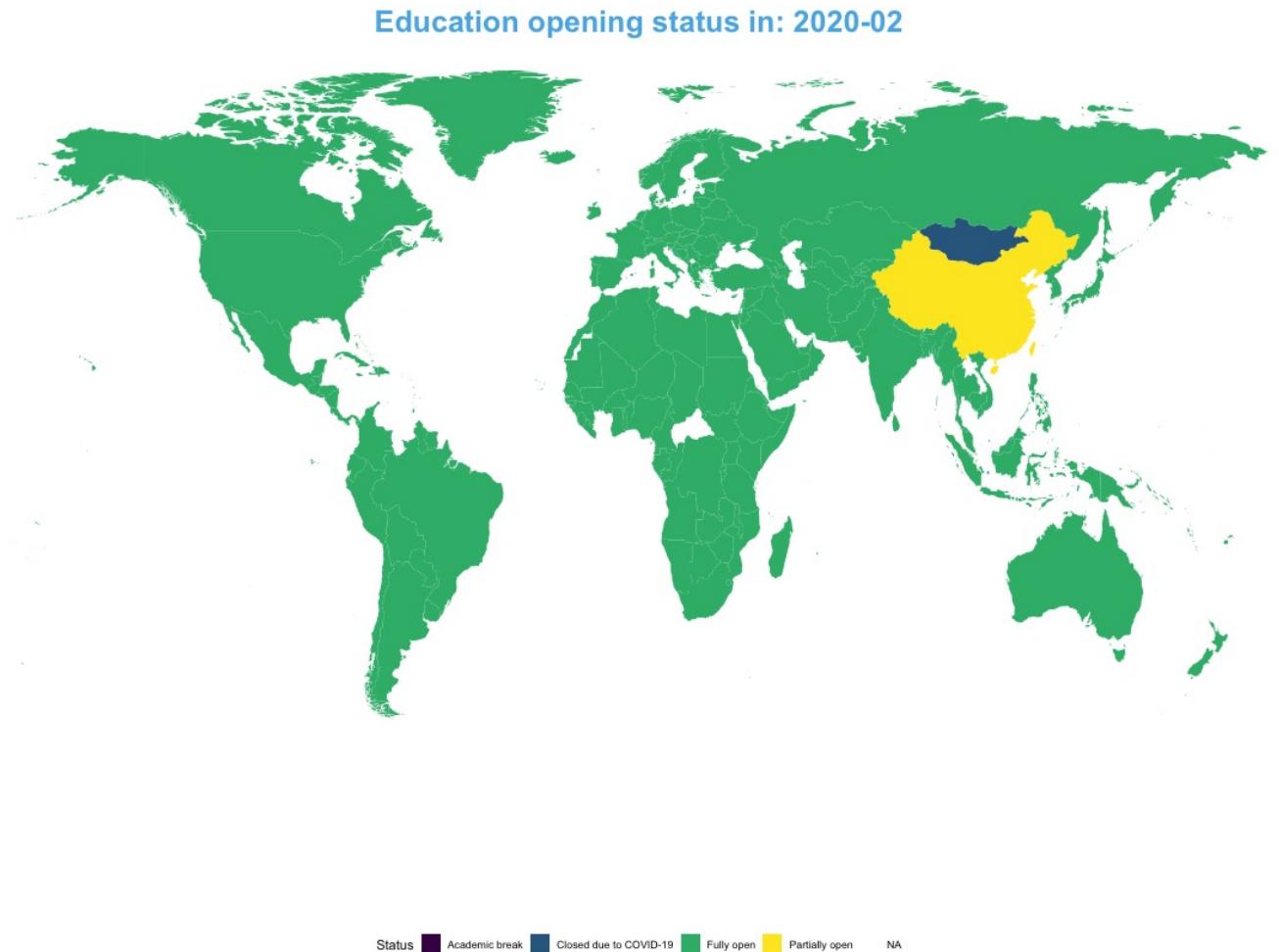


Education Status VS Total Cases and Pop % in Most Severe Countries in Oct





- Status of educational institutes worldwide from 2020 Feb to 2020 Oct (monthly)



Top 10 Countries Ranking with Most Cumulated Covid-19 Cases: 2020-01

China | 1274

Thailand | 6

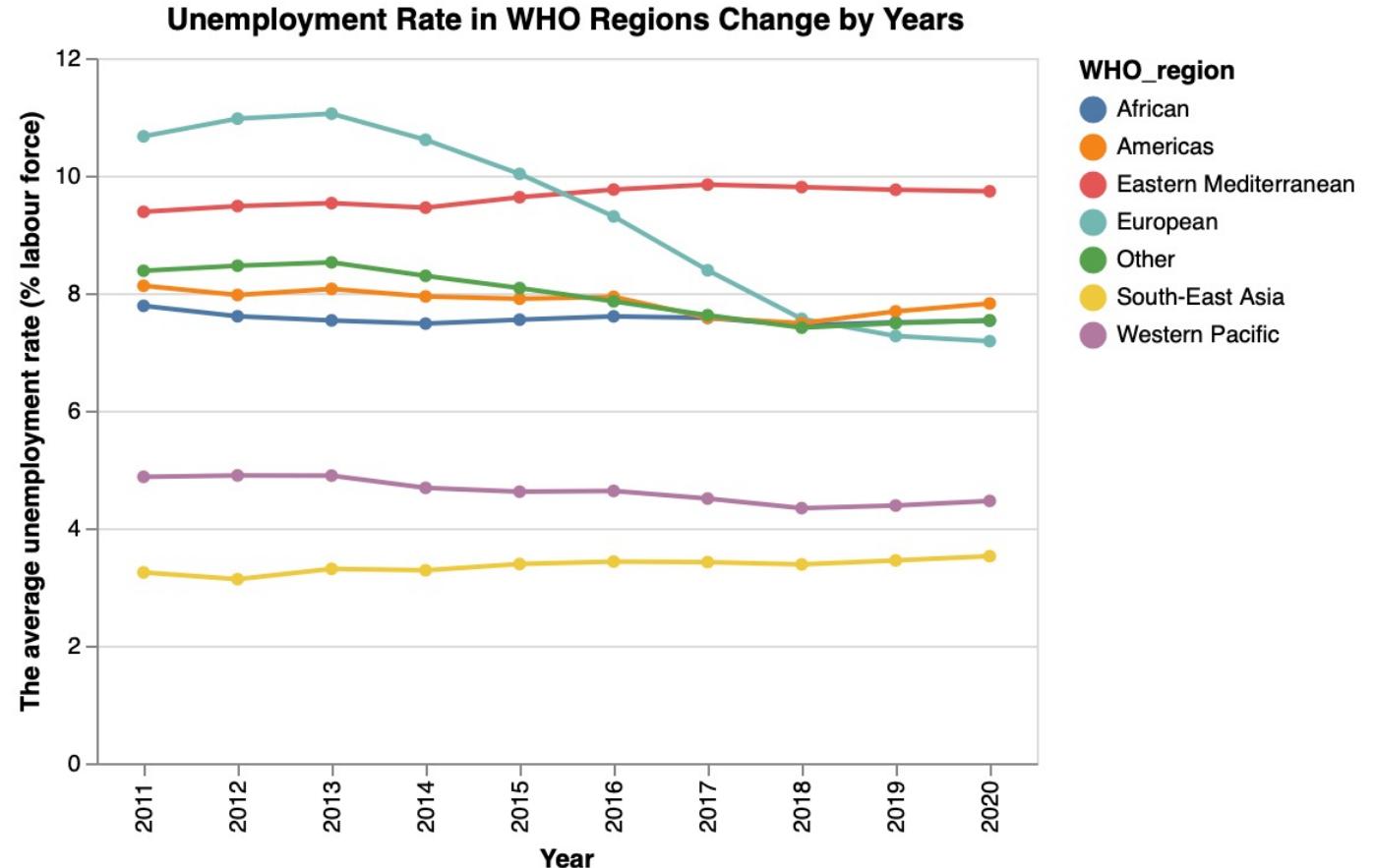
United Republic of Korea | 3

Wales | 2

- Top 10 countries with the most cumulative confirmed cases from 2020 Jan to present



- Unemployment rates changes in the last decade(by region defined by WHO)
- The result is unexpected



Legality and Ethics

WHO: Terms and Condition of Data Policy

<https://www.who.int/about/who-we-are/publishing-policies/data-policy/terms-and-conditions>

World Bank: Terms of Use for Datasets

<https://www.worldbank.org/en/about/legal/terms-of-use-for-datasets>

Break the law

Yes No

Unethical

Yes No

Harmful to any person or institution

Yes No

Intended Use

Used as the basis for many different directions of research

- ✓ *Create a time series model to predict future changes of Covid-19*
- ✓ *Forecasting economic and educational trends*
- ✓ *Assessing the impact of Covid-19 on the future economy*