INSTALLING PACKAGES

```
#installing packages
install.packages("httr")
install.packages("rvest")
```

```
    Installing package into ‘/usr/local/lib/R/site-library’
    (as ‘lib’ is unspecified)

    Installing package into ‘/usr/local/lib/R/site-library’
    (as ‘lib’ is unspecified)
```

LOADING LIBRARIES

```
#loading libraries
library(httr)
library(rvest)
```

TASK 1: Get a COVID-19 pandemic Wiki page using HTTP request

```
# Specify the URL of the webpage
url <- "https://en.wikipedia.org/w/index.php?title=Template:COVID-19_test

# Read the HTML content of the webpage
page <- read_html(url)
```

TASK 2: Extract COVID-19 testing data table from the wiki HTML page

```
# Example: Scrape the table data
table <- page %>% html_node("table.wikitable") %>% html_table()
```

DATA CHECKS

```
head(table)
```

A tibble: 6 ×

Country or region	Date[a]	Tested	Units[b]	Confirmed(cases)	Confirmed /te
<chr>	<chr>	<chr>	<chr>	<chr>	
Afghanistan	17 Dec 2020	154,767	samples	49,621	
Albania	18 Feb 2021	428,654	samples	96,838	
Algeria	2 Nov 2020	230,553	samples	58,574	
Andorra	23 Feb 2022	300,307	samples	37,958	
Angola	2 Feb 2021	399,228	samples	20,981	
Antigua and Barbuda	6 Mar 2021	15,268	samples	832	

```
tail(table)
```

Country or region	Date[a]	Tested	Units[b]	Conf

region			
<chr>	<chr>	<chr>	<chr>
Uzbekistan	7 Sep 2020	2,630,000	samples
Venezuela	30 Mar 2021	3,179,074	samples
Vietnam	28 Aug 2022	45,772,571	samples
Zambia	10 Mar 2022	3,301,860	samples
Zimbabwe	15 Oct 2022	2,529,087	samples
.mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;list-style-type:decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bottom:0;list-style-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30em}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-columns{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .reflist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mw-parser-output .reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .reflist-	.mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;list-style-type:decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bottom:0;list-style-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30em}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-columns{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .reflist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mw-parser-output .reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .reflist-	.mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;list-style-type:decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bottom:0;list-style-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30em}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-columns{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .reflist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mw-parser-output .reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .reflist-	.mw-parser-output .reflist{font-size:90%;margin-bottom:0.5em;list-style-type:decimal}.mw-parser-output .reflist .references{font-size:100%;margin-bottom:0;list-style-type:inherit}.mw-parser-output .reflist-columns-2{column-width:30em}.mw-parser-output .reflist-columns-3{column-width:25em}.mw-parser-output .reflist-columns{margin-top:0.3em}.mw-parser-output .reflist-columns ol{margin-top:0}.mw-parser-output .reflist-columns li{page-break-inside:avoid;break-inside:avoid-column}.mw-parser-output .reflist-upper-alpha{list-style-type:upper-alpha}.mw-parser-output .reflist-

```
output .reflist-upper-roman{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha{list-style-type:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type:lower-greek}.mw-parser-output .reflist-lower-roman{list-style-type:lower-roman}
```

^ Local time. ^ For some countries it is unclear whether they report samples or cases. One person tested twice is recorded as one case and two samples. ^ Excluding Taiwan. ^ Excluding Northern Cyprus. ^ Excluding Greenland and the Faroe Islands. ^ Excluding Overseas France. ^ Testing data from 4 May to 12 May is missing because of the transition to the new reporting system SI-DEP. ^ Excluding Abkhazia and South Ossetia. ^ Data for residents only. ^ Excluding Transnistria. ^ Northern Cyprus is not recognized as a sovereign state by any country except Turkey. ^

output .reflist-upper-roman{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha{list-style-type:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type:lower-greek}.mw-parser-output .reflist-lower-roman{list-style-type:lower-roman} ^ Local time. ^ For some countries it is unclear whether they report samples or cases. One person tested twice is recorded as one case and two samples. ^ Excluding Taiwan. ^ Excluding Northern Cyprus. ^ Excluding Greenland and the Faroe Islands. ^ Excluding Overseas France. ^ Testing data from 4 May to 12 May is missing because of the transition to the new reporting system SI-DEP. ^ Excluding Abkhazia and South Ossetia. ^ Data for residents only. ^ Excluding Transnistria. ^ Northern Cyprus is not recognized as a sovereign state by any country except Turkey. ^

output .reflist-upper-roman{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha{list-style-type:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type:lower-greek}.mw-parser-output .reflist-lower-roman{list-style-type:lower-roman} ^ Local time. ^ For some countries it is unclear whether they report samples or cases. One person tested twice is recorded as one case and two samples. ^ Excluding Taiwan. ^ Excluding Northern Cyprus. ^ Excluding Greenland and the Faroe Islands. ^ Excluding Overseas France. ^ Testing data from 4 May to 12 May is missing because of the transition to the new reporting system SI-DEP. ^ Excluding Abkhazia and South Ossetia. ^ Data for residents only. ^ Excluding Transnistria. ^ Northern Cyprus is not recognized as a sovereign state by any country except Turkey. ^

output .reflist-upper-roman{list-style-type:upper-roman}.mw-parser-output .reflist-lower-alpha{list-style-type:lower-alpha}.mw-parser-output .reflist-lower-greek{list-style-type:lower-greek}.mw-parser-output .reflist-lower-roman{list-style-type:lower-roman} ^ Local time. ^ For some countries it is unclear whether they report samples or cases. One person tested twice is recorded as one case and two samples. ^ Excluding Taiwan. ^ Excluding Northern Cyprus. ^ Excluding Greenland and the Faroe Islands. ^ Excluding Overseas France. ^ Testing data from 4 May to 12 May is missing because of the transition to the new reporting system SI-DEP. ^ Excluding Abkhazia and South Ossetia. ^ Data for residents only. ^ Excluding Transnistria. ^ Northern Cyprus is not recognized as a sovereign state by any country except Turkey. ^

Includes data for Liechtenstein. ^ Not a United Nations member.	Includes data for Liechtenstein. ^ Not a United Nations member.	Includes data for Liechtenstein. ^ Not a United Nations member.	Includes data for Liechtenstein. ^ Not a United Nations member.

```
str(table)
```

```
tibble [173 × 9] (S3: tbl_df/tbl/data.frame)
 $ Country or region      : chr [1:173] "Afghanistan" "Albania" "Alg
 $ Date[a]                : chr [1:173] "17 Dec 2020" "18 Feb 2021"
 $ Tested                 : chr [1:173] "154,767" "428,654" "230,553
 $ Units[b]               : chr [1:173] "samples" "samples" "samples
 $ Confirmed(cases)       : chr [1:173] "49,621" "96,838" "58,574" "
 $ Confirmed /tested,%    : chr [1:173] "32.1" "22.6" "25.4" "12.6"
 $ Tested /population,%   : chr [1:173] "0.40" "15.0" "0.53" "387" .
 $ Confirmed /population,%: chr [1:173] "0.13" "3.4" "0.13" "49.0" .
 $ Ref.                   : chr [1:173] "[1]" "[2]" "[3][4]" "[5]" .
```

```
# Check the shape of the dataframe
shape <- dim(table)
print(shape)
```

```
[1] 173   9
```

TASK 3:PRE PROCESSING

```
# Remove the last row
table <- table[1:(nrow(table)-1), ]
table
```

A tibble: 1

Country or region	Date[a]	Tested	Units[b]	Confirmed(cases)	Confirme
<chr>	<chr>	<chr>	<chr>	<chr>	
Afghanistan	17 Dec 2020	154,767	samples	49,621	
Albania	18 Feb 2021	428,654	samples	96,838	
Algeria	2 Nov 2020	230,553	samples	58,574	

Andorra	23 Feb 2022	300,307	samples	37,958
Angola	2 Feb 2021	399,228	samples	20,981
Antigua and Barbuda	6 Mar 2021	15,268	samples	832
Argentina	16 Apr 2022	35,716,069	samples	9,060,495
Armenia	29 May 2022	3,099,602	samples	422,963
Australia	9 Sep 2022	78,548,492	samples	10,112,229
Austria	1 Feb 2023	205,817,752	samples	5,789,991
Azerbaijan	11 May 2022	6,838,458	samples	792,638
Bahamas	28 Nov 2022	259,366	samples	37,483
Bahrain	3 Dec 2022	10,578,766	samples	696,614
Bangladesh	24 Jul 2021	7,417,714	samples	1,151,644
Barbados	14 Oct 2022	770,100	samples	103,014
Belarus	9 May 2022	13,217,569	samples	982,809
Belgium	24 Jan 2023	36,548,544	samples	4,691,499
Belize	8 Jun 2022	572,900	samples	60,694
Benin	4 May 2021	595,112	samples	7,884
Bhutan	28 Feb 2022	1,736,168	samples	12,702
Bolivia	5 Jun 2022	4,358,669	cases	910,228
Bosnia and	27 Sep			

Bosnia and Herzegovina	27 Sep 2022	1,872,934	samples	399,887
Botswana	11 Jan 2022	2,026,898		232,432
Brazil	19 Feb 2021	23,561,497	samples	10,081,676
Brunei	2 Aug 2021	153,804	samples	338
Bulgaria	2 Feb 2023	10,993,239	samples	1,295,524
Burkina Faso	4 Mar 2021	158,777	samples	12,123
Burundi	5 Jan 2021	90,019		884
Cambodia	1 Aug 2021	1,812,706		77,914
Cameroon	18 Feb 2021	942,685	samples	32,681
⋮	⋮	⋮	⋮	⋮
Serbia	2 Feb 2023	12,185,475	cases	2,473,599
Singapore	3 Aug 2021	16,206,203	samples	65,315
Slovakia	2 Feb 2023	7,391,882	samples	1,861,034
Slovenia	2 Feb 2023	2,826,117	samples	1,322,282
South Africa	24 May 2021	11,378,282	cases	1,637,848
South Korea	1 Mar 2021	6,592,010	samples	90,029
South Sudan	26 May 2021	164,472		10,688
Spain	1 Jul 2021	54,128,524	samples	3,821,305
Sri Lanka	30 Mar 2021	2,384,745	samples	93,128

Sudan	7 Jan 2021	158,804	samples	23,316
Sweden	24 May 2021	9,996,795	samples	1,074,751
Switzerland[l]	7 Nov 2022	23,283,909	samples	4,276,836
Taiwan[m]	3 Feb 2023	30,275,725	samples	8,622,129
Tanzania	18 Nov 2020	3,880		509
Thailand	4 Mar 2021	1,579,597	cases	26,162
Togo	6 Jan 2023	807,269		39,358
Trinidad and Tobago	3 Jan 2022	512,730	cases	92,997
Tunisia	23 Aug 2021	2,893,625	samples	703,732
Turkey	2 Jul 2021	61,236,294	samples	5,435,831
Uganda	11 Feb 2021	852,444	samples	39,979
Ukraine	24 Nov 2021	15,648,456	samples	3,367,461
United Arab Emirates	1 Feb 2023	198,685,717	samples	1,049,537
United Kingdom	19 May 2022	522,526,476	samples	22,232,377
United States	29 Jul 2022	929,349,291	samples	90,749,469
Uruguay	16 Apr 2022	6,089,116	samples	895,592
Uzbekistan	7 Sep 2020	2,630,000	samples	43,975
Venezuela	30 Mar 2021	3,179,074	samples	159,149
	28 Aug			

Vietnam	2022	45,772,571	samples	11,403,302
Zambia	10 Mar 2022	3,301,860	samples	314,850
Zimbabwe	15 Oct 2022	2,529,087	samples	257,893

```
# Remove the last column
table1 <- table[, -ncol(table)]
table1
```

A tibble: 172 ×

Country or region	Date[a]	Tested	Units[b]	Confirmed(cases)	Confirme
<chr>	<chr>	<chr>	<chr>	<chr>	
Afghanistan	17 Dec 2020	154,767	samples	49,621	
Albania	18 Feb 2021	428,654	samples	96,838	
Algeria	2 Nov 2020	230,553	samples	58,574	
Andorra	23 Feb 2022	300,307	samples	37,958	
Angola	2 Feb 2021	399,228	samples	20,981	
Antigua and Barbuda	6 Mar 2021	15,268	samples	832	
Argentina	16 Apr 2022	35,716,069	samples	9,060,495	
Armenia	29 May 2022	3,099,602	samples	422,963	
Australia	9 Sep 2022	78,548,492	samples	10,112,229	
Austria	1 Feb 2023	205,817,752	samples	5,789,991	
Azerbaijan	11 May 2022	6,838,458	samples	792,638	

Bahamas	28 Nov 2022	259,366	samples	37,483
Bahrain	3 Dec 2022	10,578,766	samples	696,614
Bangladesh	24 Jul 2021	7,417,714	samples	1,151,644
Barbados	14 Oct 2022	770,100	samples	103,014
Belarus	9 May 2022	13,217,569	samples	982,809
Belgium	24 Jan 2023	36,548,544	samples	4,691,499
Belize	8 Jun 2022	572,900	samples	60,694
Benin	4 May 2021	595,112	samples	7,884
Bhutan	28 Feb 2022	1,736,168	samples	12,702
Bolivia	5 Jun 2022	4,358,669	cases	910,228
Bosnia and Herzegovina	27 Sep 2022	1,872,934	samples	399,887
Botswana	11 Jan 2022	2,026,898		232,432
Brazil	19 Feb 2021	23,561,497	samples	10,081,676
Brunei	2 Aug 2021	153,804	samples	338
Bulgaria	2 Feb 2023	10,993,239	samples	1,295,524
Burkina Faso	4 Mar 2021	158,777	samples	12,123
Burundi	5 Jan 2021	90,019		884
Cambodia	1 Aug 2021	1,812,706		77,914
	18 Feb			

Cameroon	2021	942,685	samples	32,681
⋮	⋮	⋮	⋮	⋮
Serbia	2 Feb 2023	12,185,475	cases	2,473,599
Singapore	3 Aug 2021	16,206,203	samples	65,315
Slovakia	2 Feb 2023	7,391,882	samples	1,861,034
Slovenia	2 Feb 2023	2,826,117	samples	1,322,282
South Africa	24 May 2021	11,378,282	cases	1,637,848
South Korea	1 Mar 2021	6,592,010	samples	90,029
South Sudan	26 May 2021	164,472		10,688
Spain	1 Jul 2021	54,128,524	samples	3,821,305
Sri Lanka	30 Mar 2021	2,384,745	samples	93,128
Sudan	7 Jan 2021	158,804	samples	23,316
Sweden	24 May 2021	9,996,795	samples	1,074,751
Switzerland[l]	7 Nov 2022	23,283,909	samples	4,276,836
Taiwan[m]	3 Feb 2023	30,275,725	samples	8,622,129
Tanzania	18 Nov 2020	3,880		509
Thailand	4 Mar 2021	1,579,597	cases	26,162
Togo	6 Jan 2023	807,269		39,358
Trinidad and Tobago	3 Jan 2022	512,730	cases	92,997

Tunisia	23 Aug 2021	2,893,625	samples	703,732
Turkey	2 Jul 2021	61,236,294	samples	5,435,831
Uganda	11 Feb 2021	852,444	samples	39,979
Ukraine	24 Nov 2021	15,648,456	samples	3,367,461
United Arab Emirates	1 Feb 2023	198,685,717	samples	1,049,537
United Kingdom	19 May 2022	522,526,476	samples	22,232,377
United States	29 Jul 2022	929,349,291	samples	90,749,469
Uruguay	16 Apr 2022	6,089,116	samples	895,592
Uzbekistan	7 Sep 2020	2,630,000	samples	43,975
Venezuela	30 Mar 2021	3,179,074	samples	159,149
Vietnam	28 Aug 2022	45,772,571	samples	11,403,302
Zambia	10 Mar 2022	3,301,860	samples	314,850
Zimbabwe	15 Oct 2022	2,529,087	samples	257,893

```
# Remove the 4th column
table1 <- table1[, -4]
```

```
# Rename the columns
colnames(table1) <- c("country", "date","tested", "confirmed", "confirmed
table1
```

A tibble: 172 × 7

country	date	tested	confirmed	confirmed.tested ratio	tested.po

<chr>	<chr>	<chr>	<chr>	<chr>
Afghanistan	17 Dec 2020	154,767	49,621	32.1
Albania	18 Feb 2021	428,654	96,838	22.6
Algeria	2 Nov 2020	230,553	58,574	25.4
Andorra	23 Feb 2022	300,307	37,958	12.6
Angola	2 Feb 2021	399,228	20,981	5.3
Antigua and Barbuda	6 Mar 2021	15,268	832	5.4
Argentina	16 Apr 2022	35,716,069	9,060,495	25.4
Armenia	29 May 2022	3,099,602	422,963	13.6
Australia	9 Sep 2022	78,548,492	10,112,229	12.9
Austria	1 Feb 2023	205,817,752	5,789,991	2.8
Azerbaijan	11 May 2022	6,838,458	792,638	11.6
Bahamas	28 Nov 2022	259,366	37,483	14.5
Bahrain	3 Dec 2022	10,578,766	696,614	6.6
Bangladesh	24 Jul 2021	7,417,714	1,151,644	15.5
Barbados	14 Oct 2022	770,100	103,014	13.4
Belarus	9 May	13,217,569	982,809	7.4

	2022			
Belgium	24 Jan 2023	36,548,544	4,691,499	12.8
Belize	8 Jun 2022	572,900	60,694	10.6
Benin	4 May 2021	595,112	7,884	1.3
Bhutan	28 Feb 2022	1,736,168	12,702	0.73
Bolivia	5 Jun 2022	4,358,669	910,228	20.9
Bosnia and Herzegovina	27 Sep 2022	1,872,934	399,887	21.4
Botswana	11 Jan 2022	2,026,898	232,432	11.5
Brazil	19 Feb 2021	23,561,497	10,081,676	42.8
Brunei	2 Aug 2021	153,804	338	0.22
Bulgaria	2 Feb 2023	10,993,239	1,295,524	11.8
Burkina Faso	4 Mar 2021	158,777	12,123	7.6
Burundi	5 Jan 2021	90,019	884	0.98
Cambodia	1 Aug 2021	1,812,706	77,914	4.3
Cameroon	18 Feb 2021	942,685	32,681	3.5
⋮	⋮	⋮	⋮	⋮
Serbia	2 Feb 2023	12,185,475	2,473,599	20.3
Singapore	3 Aug	16,206,203	65,315	0.40

Singapore	2021	10,200,203	65,515	0.40
Slovakia	2 Feb 2023	7,391,882	1,861,034	25.2
Slovenia	2 Feb 2023	2,826,117	1,322,282	46.8
South Africa	24 May 2021	11,378,282	1,637,848	14.4
South Korea	1 Mar 2021	6,592,010	90,029	1.4
South Sudan	26 May 2021	164,472	10,688	6.5
Spain	1 Jul 2021	54,128,524	3,821,305	7.1
Sri Lanka	30 Mar 2021	2,384,745	93,128	3.9
Sudan	7 Jan 2021	158,804	23,316	14.7
Sweden	24 May 2021	9,996,795	1,074,751	10.8
Switzerland[l]	7 Nov 2022	23,283,909	4,276,836	18.4
Taiwan[m]	3 Feb 2023	30,275,725	8,622,129	28.48
Tanzania	18 Nov 2020	3,880	509	13.1
Thailand	4 Mar 2021	1,579,597	26,162	1.7
Togo	6 Jan 2023	807,269	39,358	4.9
Trinidad and Tobago	3 Jan 2022	512,730	92,997	18.1
Tunisia	23 Aug 2021	2,893,625	703,732	24.3

Turkey	2 Jul 2021	61,236,294	5,435,831	8.9
Uganda	11 Feb 2021	852,444	39,979	4.7
Ukraine	24 Nov 2021	15,648,456	3,367,461	21.5
United Arab Emirates	1 Feb 2023	198,685,717	1,049,537	0.53
United Kingdom	19 May 2022	522,526,476	22,232,377	4.3
United States	29 Jul 2022	929,349,291	90,749,469	9.8
Uruguay	16 Apr 2022	6,089,116	895,592	14.7
Uzbekistan	7 Sep 2020	2,630,000	43,975	1.7
Venezuela	30 Mar 2021	3,179,074	159,149	5.0
Vietnam	28 Aug 2022	45,772,571	11,403,302	24.9
Zambia	10 Mar 2022	3,301,860	314,850	9.5
Zimbabwe	15 Oct 2022	2,529,087	257,893	10.2

```
str(table1)
```

```
tibble [172 × 7] (S3: tbl_df/tbl/data.frame)
 $ country                   : chr [1:172] "Afghanistan" "Albania" '
 $ date                      : chr [1:172] "17 Dec 2020" "18 Feb 202
 $ tested                    : chr [1:172] "154,767" "428,654" "230,
 $ confirmed                 : chr [1:172] "49,621" "96,838" "58,574
 $ confirmed.tested ratio    : chr [1:172] "32.1" "22.6" "25.4" "12.
 $ tested.population.ratio   : chr [1:172] "0.40" "15.0" "0.53" "387
 $ confirmed.population.ratio: chr [1:172] "0.13" "3.4" "0.13" "49.0
```

```
# Check the shape of the dataframe
shape <- dim(table1)
print(shape)
```

```
[1] 172   7
```

```
# Remove commas from numeric columns and convert to numeric data type
table1$tested <- as.numeric(gsub(",", "", table1$tested))
table1$`confirmed` <- as.numeric(gsub(",", "", table1$`confirmed`))
table1
```

A tibble: 172 × 7

country <chr>	date <chr>	tested <dbl>	confirmed <dbl>	confirmed.tested ratio <chr>	tested.popu
Afghanistan	17 Dec 2020	154767	49621	32.1	
Albania	18 Feb 2021	428654	96838	22.6	
Algeria	2 Nov 2020	230553	58574	25.4	
Andorra	23 Feb 2022	300307	37958	12.6	
Angola	2 Feb 2021	399228	20981	5.3	
Antigua and Barbuda	6 Mar 2021	15268	832	5.4	

Argentina	16 Apr 2022	35716069	9060495	25.4
Armenia	29 May 2022	3099602	422963	13.6
Australia	9 Sep 2022	78548492	10112229	12.9
Austria	1 Feb 2023	205817752	5789991	2.8
Azerbaijan	11 May 2022	6838458	792638	11.6
Bahamas	28 Nov 2022	259366	37483	14.5
Bahrain	3 Dec 2022	10578766	696614	6.6
Bangladesh	24 Jul 2021	7417714	1151644	15.5
Barbados	14 Oct 2022	770100	103014	13.4
Belarus	9 May 2022	13217569	982809	7.4
Belgium	24 Jan 2023	36548544	4691499	12.8
Belize	8 Jun 2022	572900	60694	10.6
Benin	4 May 2021	595112	7884	1.3
Bhutan	28 Feb 2022	1736168	12702	0.73
Bolivia	5 Jun 2022	4358669	910228	20.9
Bosnia and Herzegovina	27 Sep 2022	1872934	399887	21.4

Botswana	11 Jan 2022	2026898	232432	11.5
Brazil	19 Feb 2021	23561497	10081676	42.8
Brunei	2 Aug 2021	153804	338	0.22
Bulgaria	2 Feb 2023	10993239	1295524	11.8
Burkina Faso	4 Mar 2021	158777	12123	7.6
Burundi	5 Jan 2021	90019	884	0.98
Cambodia	1 Aug 2021	1812706	77914	4.3
Cameroon	18 Feb 2021	942685	32681	3.5
⋮	⋮	⋮	⋮	⋮
Serbia	2 Feb 2023	12185475	2473599	20.3
Singapore	3 Aug 2021	16206203	65315	0.40
Slovakia	2 Feb 2023	7391882	1861034	25.2
Slovenia	2 Feb 2023	2826117	1322282	46.8
South Africa	24 May 2021	11378282	1637848	14.4
South Korea	1 Mar 2021	6592010	90029	1.4
South Sudan	26 May 2021	164472	10688	6.5
Spain	1 Jul 2021	54128524	3821305	7.1
	30			

Sri Lanka	Mar 2021	2384745	93128	3.9
Sudan	7 Jan 2021	158804	23316	14.7
Sweden	24 May 2021	9996795	1074751	10.8
Switzerland[l]	7 Nov 2022	23283909	4276836	18.4
Taiwan[m]	3 Feb 2023	30275725	8622129	28.48
Tanzania	18 Nov 2020	3880	509	13.1
Thailand	4 Mar 2021	1579597	26162	1.7
Togo	6 Jan 2023	807269	39358	4.9
Trinidad and Tobago	3 Jan 2022	512730	92997	18.1
Tunisia	23 Aug 2021	2893625	703732	24.3
Turkey	2 Jul 2021	61236294	5435831	8.9
Uganda	11 Feb 2021	852444	39979	4.7
Ukraine	24 Nov 2021	15648456	3367461	21.5
United Arab Emirates	1 Feb 2023	198685717	1049537	0.53
United Kingdom	19 May 2022	522526476	22232377	4.3
United States	29 Jul 2022	929349291	90749469	9.8

Uruguay	16 Apr 2022	6089116	895592	14.7
Uzbekistan	7 Sep 2020	2630000	43975	1.7
Venezuela	30 Mar 2021	3179074	159149	5.0
Vietnam	28 Aug 2022	45772571	11403302	24.9
Zambia	10 Mar 2022	3301860	314850	9.5
Zimbabwe	15 Oct 2022	2529087	257893	10.2

SAVING DATAFRAME

```
# Save the dataframe to a CSV file
write.csv(table1, "table1_data.csv", row.names = FALSE)

# Print a confirmation message
cat("Data saved as 'table1_data.csv'")
```

```
Data saved as 'table1_data.csv'
```

ANALYSIS

TASK 4: Get a subset of the extracted data frame

The goal of task 4 is to get the 5th to 10th rows from the data frame with only `country` and `confirmed` columns selected

```
# Read covid_data_frame_csv from the csv file
```

```
table1
```

A tibble: 172 × 7

country <chr>	date <chr>	tested <dbl>	confirmed <dbl>	confirmed.tested ratio <chr>	tested.popu
Afghanistan	17 Dec 2020	154767	49621	32.1	
Albania	18 Feb 2021	428654	96838	22.6	
Algeria	2 Nov 2020	230553	58574	25.4	
Andorra	23 Feb 2022	300307	37958	12.6	
Angola	2 Feb 2021	399228	20981	5.3	
Antigua and Barbuda	6 Mar 2021	15268	832	5.4	
Argentina	16 Apr 2022	35716069	9060495	25.4	
Armenia	29 May 2022	3099602	422963	13.6	
Australia	9 Sep 2022	78548492	10112229	12.9	
Austria	1 Feb 2023	205817752	5789991	2.8	
Azerbaijan	11 May 2022	6838458	792638	11.6	
Bahamas	28 Nov 2022	259366	37483	14.5	
Bahrain	3 Dec 2022	10578766	696614	6.6	
Bangladesh	24 Jul	7417714	1151944	15.5	

Bangladesh	2021	7417714	1151644	15.5
Barbados	14 Oct 2022	770100	103014	13.4
Belarus	9 May 2022	13217569	982809	7.4
Belgium	24 Jan 2023	36548544	4691499	12.8
Belize	8 Jun 2022	572900	60694	10.6
Benin	4 May 2021	595112	7884	1.3
Bhutan	28 Feb 2022	1736168	12702	0.73
Bolivia	5 Jun 2022	4358669	910228	20.9
Bosnia and Herzegovina	27 Sep 2022	1872934	399887	21.4
Botswana	11 Jan 2022	2026898	232432	11.5
Brazil	19 Feb 2021	23561497	10081676	42.8
Brunei	2 Aug 2021	153804	338	0.22
Bulgaria	2 Feb 2023	10993239	1295524	11.8
Burkina Faso	4 Mar 2021	158777	12123	7.6
Burundi	5 Jan 2021	90019	884	0.98
Cambodia	1 Aug 2021	1812706	77914	4.3
Cameroon	18 Feb 2021	942685	32681	3.5

Serbia	2 Feb 2023	12185475	2473599	20.3
Singapore	3 Aug 2021	16206203	65315	0.40
Slovakia	2 Feb 2023	7391882	1861034	25.2
Slovenia	2 Feb 2023	2826117	1322282	46.8
South Africa	24 May 2021	11378282	1637848	14.4
South Korea	1 Mar 2021	6592010	90029	1.4
South Sudan	26 May 2021	164472	10688	6.5
Spain	1 Jul 2021	54128524	3821305	7.1
Sri Lanka	30 Mar 2021	2384745	93128	3.9
Sudan	7 Jan 2021	158804	23316	14.7
Sweden	24 May 2021	9996795	1074751	10.8
Switzerland[l]	7 Nov 2022	23283909	4276836	18.4
Taiwan[m]	3 Feb 2023	30275725	8622129	28.48
Tanzania	18 Nov 2020	3880	509	13.1
Thailand	4 Mar 2021	1579597	26162	1.7
Togo	6 Jan 2023	807269	39358	4.9

Trinidad and Tobago	3 Jan 2022	512730	92997	18.1
Tunisia	23 Aug 2021	2893625	703732	24.3
Turkey	2 Jul 2021	61236294	5435831	8.9
Uganda	11 Feb 2021	852444	39979	4.7
Ukraine	24 Nov 2021	15648456	3367461	21.5
United Arab Emirates	1 Feb 2023	198685717	1049537	0.53
United Kingdom	19 May 2022	522526476	22232377	4.3
United States	29 Jul 2022	929349291	90749469	9.8
Uruguay	16 Apr 2022	6089116	895592	14.7
Uzbekistan	7 Sep 2020	2630000	43975	1.7
Venezuela	30 Mar 2021	3179074	159149	5.0
Vietnam	28 Aug 2022	45772571	11403302	24.9
Zambia	10 Mar 2022	3301860	314850	9.5
Zimbabwe	15 Oct 2022	2529087	257893	10.2

```
# Get the 5th to 10th rows, with two "country" "confirmed"

subset_table <- table1[5:10, c("country", "confirmed", "country", "confir

# Print the subset table
print(subset_table)
```

```
# A tibble: 6 × 4
  country             confirmed country             confirmed
  <chr>                   <dbl> <chr>                   <dbl>
1 Angola                  20981 Angola                  20981
2 Antigua and Barbuda       832 Antigua and Barbuda       832
3 Argentina             9060495 Argentina             9060495
4 Armenia                422963 Armenia                422963
5 Australia            10112229 Australia            10112229
6 Austria               5789991 Austria               5789991
```

TASK 5: Calculate worldwide COVID testing positive ratio

The goal of task 5 is to get the total confirmed and tested cases worldwide, and try to figure the overall positive ratio using confirmed cases / tested cases

```
# Calculate the total confirmed cases worldwide
total_confirmed <- sum(table1$`confirmed`, na.rm = TRUE)
total_confirmed
```

```
431430484
```

```
# Calculate the total tested cases worldwide
total_tested <- sum(table1$tested, na.rm = TRUE)
total_tested
```

```
5396633953
```

```
# Calculate the positive ratio (confirmed / tested)
positive_ratio <- total_confirmed / total_tested
positive_ratio
```

```
0.0799443667584989
```

TASK 6: Get a country list which reported their testing data

The goal of task 6 is to get a catalog or sorted list of countries who have reported their COVID-19 testing data

```
# Get the country column
country_column <- table1$country
country_column
```

```
'Afghanistan' · 'Albania' · 'Algeria' · 'Andorra' · 'Angola' · 'Antigua and Barbuda' · 'Argentina' · 'Armenia' · 'Australia' · 'Austria' · 'Azerbaijan' · 'Bahamas' · 'Bahrain' · 'Bangladesh' · 'Barbados' · 'Belarus' · 'Belgium' · 'Belize' · 'Benin' · 'Bhutan' · 'Bolivia' · 'Bosnia and Herzegovina' · 'Botswana' · 'Brazil' · 'Brunei' · 'Bulgaria' · 'Burkina Faso' · 'Burundi' · 'Cambodia' · 'Cameroon' · 'Canada' · 'Chad' · 'Chile' · 'China[c]' · 'Colombia' · 'Costa Rica' · 'Croatia' · 'Cuba' · 'Cyprus[d]' · 'Czechia' · 'Denmark[e]' · 'Djibouti' · 'Dominica' · 'Dominican Republic' · 'DR Congo' · 'Ecuador' · 'Egypt' · 'El Salvador' · 'Equatorial Guinea' · 'Estonia' · 'Eswatini' · 'Ethiopia' · 'Faroe Islands' · 'Fiji' · 'Finland' · 'France[f][g]' · 'Gabon' · 'Gambia' · 'Georgia[h]' · 'Germany' · 'Ghana' · 'Greece' · 'Greenland' · 'Grenada' · 'Guatemala' · 'Guinea' · 'Guinea-Bissau' · 'Guyana' · 'Haiti' · 'Honduras' · 'Hungary' · 'Iceland' · 'India' · 'Indonesia' · 'Iran' · 'Iraq' · 'Ireland' · 'Israel' · 'Italy' · 'Ivory Coast' · 'Jamaica' · 'Japan' · 'Jordan' · 'Kazakhstan' · 'Kenya' · 'Kosovo' · 'Kuwait' · 'Kyrgyzstan' · 'Laos' · 'Latvia' · 'Lebanon' · 'Lesotho' · 'Liberia' · 'Libya' · 'Lithuania' · 'Luxembourg[i]' · 'Madagascar' · 'Malawi' · 'Malaysia' · 'Maldives' · 'Mali' · 'Malta' · 'Mauritania' · 'Mauritius' · 'Mexico' · 'Moldova[j]' · 'Mongolia' · 'Montenegro' · 'Morocco' · 'Mozambique' · 'Myanmar' · 'Namibia' · 'Nepal' · 'Netherlands' · 'New Caledonia' · 'New Zealand' · 'Niger' · 'Nigeria' · 'North Korea' · 'North Macedonia' · 'Northern Cyprus[k]' · 'Norway' · 'Oman' · 'Pakistan' · 'Palestine' · 'Panama' · 'Papua New Guinea' · 'Paraguay' · 'Peru' · 'Philippines' · 'Poland' · 'Portugal' · 'Qatar' · 'Romania' · 'Russia' · 'Rwanda' · 'Saint Kitts and Nevis' · 'Saint Lucia' · 'Saint Vincent' · 'San Marino' · 'Saudi Arabia' · 'Senegal' · 'Serbia' · 'Singapore' · 'Slovakia' · 'Slovenia' · 'South Africa' · 'South Korea' · 'South Sudan' · 'Spain' · 'Sri Lanka' · 'Sudan' · 'Sweden' · 'Switzerland[l]' · 'Taiwan[m]' · 'Tanzania' · 'Thailand' · 'Togo' · 'Trinidad and Tobago' · 'Tunisia' · 'Turkey' · 'Uganda' · 'Ukraine' · 'United Arab Emirates' · 'United Kingdom' · 'United States' · 'Uruguay' · 'Uzbekistan' · 'Venezuela' · 'Vietnam' · 'Zambia' · 'Zimbabwe'
```

```
# Check the class of the country column
country_class <- class(country_column)
cat("Class of the country column:", country_class, "\n")
```

```
Class of the country column: character
```

```
# Convert the country column to character type
country_column <- as.character(country_column)
country_column
```

'Afghanistan' · 'Albania' · 'Algeria' · 'Andorra' · 'Angola' · 'Antigua and Barbuda' · 'Argentina' · 'Armenia' · 'Australia' · 'Austria' · 'Azerbaijan' · 'Bahamas' · 'Bahrain' · 'Bangladesh' · 'Barbados' · 'Belarus' · 'Belgium' · 'Belize' · 'Benin' · 'Bhutan' · 'Bolivia' · 'Bosnia and Herzegovina' · 'Botswana' · 'Brazil' · 'Brunei' · 'Bulgaria' · 'Burkina Faso' · 'Burundi' · 'Cambodia' · 'Cameroon' · 'Canada' · 'Chad' · 'Chile' · 'China[c]' · 'Colombia' · 'Costa Rica' · 'Croatia' · 'Cuba' · 'Cyprus[d]' · 'Czechia' · 'Denmark[e]' · 'Djibouti' · 'Dominica' · 'Dominican Republic' · 'DR Congo' · 'Ecuador' · 'Egypt' · 'El Salvador' · 'Equatorial Guinea' · 'Estonia' · 'Eswatini' · 'Ethiopia' · 'Faroe Islands' · 'Fiji' · 'Finland' · 'France[f][g]' · 'Gabon' · 'Gambia' · 'Georgia[h]' · 'Germany' · 'Ghana' · 'Greece' · 'Greenland' · 'Grenada' · 'Guatemala' · 'Guinea' · 'Guinea-Bissau' · 'Guyana' · 'Haiti' · 'Honduras' · 'Hungary' · 'Iceland' · 'India' · 'Indonesia' · 'Iran' · 'Iraq' · 'Ireland' · 'Israel' · 'Italy' · 'Ivory Coast' · 'Jamaica' · 'Japan' · 'Jordan' · 'Kazakhstan' · 'Kenya' · 'Kosovo' · 'Kuwait' · 'Kyrgyzstan' · 'Laos' · 'Latvia' · 'Lebanon' · 'Lesotho' · 'Liberia' · 'Libya' · 'Lithuania' · 'Luxembourg[i]' · 'Madagascar' · 'Malawi' · 'Malaysia' · 'Maldives' · 'Mali' · 'Malta' · 'Mauritania' · 'Mauritius' · 'Mexico' · 'Moldova[j]' · 'Mongolia' · 'Montenegro' · 'Morocco' · 'Mozambique' · 'Myanmar' · 'Namibia' · 'Nepal' · 'Netherlands' · 'New Caledonia' · 'New Zealand' · 'Niger' · 'Nigeria' · 'North Korea' · 'North Macedonia' · 'Northern Cyprus[k]' · 'Norway' · 'Oman' · 'Pakistan' · 'Palestine' · 'Panama' · 'Papua New Guinea' · 'Paraguay' · 'Peru' · 'Philippines' · 'Poland' · 'Portugal' · 'Qatar' · 'Romania' · 'Russia' · 'Rwanda' · 'Saint Kitts and Nevis' · 'Saint Lucia' · 'Saint Vincent' · 'San Marino' · 'Saudi Arabia' · 'Senegal' · 'Serbia' · 'Singapore' · 'Slovakia' · 'Slovenia' · 'South Africa' · 'South Korea' · 'South Sudan' · 'Spain' · 'Sri Lanka' · 'Sudan' · 'Sweden' · 'Switzerland[l]' · 'Taiwan[m]' · 'Tanzania' · 'Thailand' · 'Togo' · 'Trinidad and Tobago' · 'Tunisia' · 'Turkey' · 'Uganda' · 'Ukraine' · 'United Arab Emirates' · 'United Kingdom' · 'United States' · 'Uruguay' · 'Uzbekistan' · 'Venezuela' · 'Vietnam' · 'Zambia' · 'Zimbabwe'

```
# Sort the countries in ascending order (A to Z)
sorted_AtoZ <- sort(country_column)
sorted_AtoZ
```

'Afghanistan' · 'Albania' · 'Algeria' · 'Andorra' · 'Angola' · 'Antigua and Barbuda' · 'Argentina' · 'Armenia' · 'Australia' · 'Austria' · 'Azerbaijan' · 'Bahamas' · 'Bahrain' · 'Bangladesh' · 'Barbados' · 'Belarus' · 'Belgium' · 'Belize' · 'Benin' · 'Bhutan' · 'Bolivia' · 'Bosnia and Herzegovina' · 'Botswana' · 'Brazil' · 'Brunei' · 'Bulgaria' · 'Burkina Faso' · 'Burundi' · 'Cambodia' · 'Cameroon' · 'Canada' · 'Chad' · 'Chile' · 'China[c]' · 'Colombia' · 'Costa Rica' · 'Croatia' · 'Cuba' · 'Cyprus[d]' · 'Czechia' · 'Denmark[e]' · 'Djibouti' · 'Dominica' · 'Dominican Republic' · 'DR Congo' · 'Ecuador' · 'Egypt' · 'El Salvador' · 'Equatorial Guinea' · 'Estonia' · 'Eswatini' · 'Ethiopia' · 'Faroe Islands' · 'Fiji' · 'Finland' · 'France[f][g]' · 'Gabon' · 'Gambia' · 'Georgia[h]' · 'Germany' · 'Ghana' · 'Greece' · 'Greenland' · 'Grenada' · 'Guatemala' · 'Guinea' · 'Guinea-Bissau' · 'Guyana' · 'Haiti' · 'Honduras' · 'Hungary' · 'Iceland' · 'India' · 'Indonesia' · 'Iran' · 'Iraq' · 'Ireland' · 'Israel' · 'Italy' · 'Ivory Coast' · 'Jamaica' · 'Japan' · 'Jordan' · 'Kazakhstan' · 'Kenya' · 'Kosovo' · 'Kuwait' · 'Kyrgyzstan' · 'Laos' · 'Latvia' · 'Lebanon' · 'Lesotho' · 'Liberia' · 'Libya' · 'Lithuania' · 'Luxembourg[i]' · 'Madagascar' · 'Malawi' · 'Malaysia' · 'Maldives' · 'Mali' · 'Malta' · 'Mauritania' · 'Mauritius' · 'Mexico' · 'Moldova[j]' · 'Mongolia' · 'Montenegro' · 'Morocco' · 'Mozambique' · 'Myanmar' · 'Namibia' · 'Nepal' · 'Netherlands' · 'New Caledonia' · 'New Zealand' · 'Niger' · 'Nigeria' · 'North Korea' · 'North Macedonia' · 'Northern Cyprus[k]' · 'Norway' · 'Oman' · 'Pakistan' · 'Palestine' · 'Panama' · 'Papua New Guinea' · 'Paraguay' · 'Peru' · 'Philippines' · 'Poland' · 'Portugal' · 'Qatar' · 'Romania' · 'Russia' · 'Rwanda' · 'Saint Kitts and Nevis' · 'Saint Lucia' · 'Saint Vincent' · 'San Marino' · 'Saudi Arabia' · 'Senegal' · 'Serbia' · 'Singapore' · 'Slovakia' · 'Slovenia' · 'South Africa' · 'South Korea' · 'South Sudan' · 'Spain' · 'Sri Lanka' · 'Sudan' · 'Sweden' · 'Switzerland[l]' · 'Taiwan[m]' · 'Tanzania' · 'Thailand' · 'Togo' · 'Trinidad and Tobago' · 'Tunisia' · 'Turkey' · 'Uganda' · 'Ukraine' · 'United Arab Emirates' · 'United Kingdom' · 'United States' · 'Uruguay' · 'Uzbekistan' · 'Venezuela' · 'Vietnam' · 'Zambia' · 'Zimbabwe'

```
# Sort the countries in descending order (Z to A)
sorted_ZtoA <- sort(country_column, decreasing = TRUE)
sorted_ZtoA
```

'Zimbabwe' · 'Zambia' · 'Vietnam' · 'Venezuela' · 'Uzbekistan' · 'Uruguay' · 'United States' · 'United Kingdom' · 'United Arab Emirates' · 'Ukraine' · 'Uganda' · 'Turkey' · 'Tunisia' · 'Trinidad and Tobago' · 'Togo' · 'Thailand' · 'Tanzania' · 'Taiwan[m]' · 'Switzerland[l]' · 'Sweden' · 'Sudan' · 'Sri Lanka' · 'Spain' · 'South Sudan' · 'South Korea' · 'South Africa' · 'Slovenia' · 'Slovakia' · 'Singapore' · 'Serbia' · 'Senegal' · 'Saudi Arabia' · 'San Marino' · 'Saint Vincent' · 'Saint Lucia' · 'Saint Kitts and Nevis' · 'Rwanda' · 'Russia' · 'Romania' · 'Qatar' · 'Portugal' · 'Poland' · 'Philippines' · 'Peru' · 'Paraguay' · 'Papua New Guinea' · 'Panama' · 'Palestine' · 'Pakistan' · 'Oman' · 'Norway' · 'Northern Cyprus[k]' · 'North Macedonia' · 'North Korea' · 'Nigeria' · 'Niger' · 'New Zealand' · 'New Caledonia' · 'Netherlands' · 'Nepal' · 'Namibia' · 'Myanmar' · 'Mozambique' · 'Morocco' · 'Montenegro' · 'Mongolia' · 'Moldova[j]' · 'Mexico' · 'Mauritius' · 'Mauritania' · 'Malta' · 'Mali' · 'Maldives' · 'Malaysia' · 'Malawi' · 'Madagascar' · 'Luxembourg[i]' · 'Lithuania' · 'Libya' · 'Liberia' · 'Lesotho' · 'Lebanon' · 'Latvia' · 'Laos' · 'Kyrgyzstan' · 'Kuwait' · 'Kosovo' · 'Kenya' · 'Kazakhstan' · 'Jordan' · 'Japan' · 'Jamaica' · 'Ivory Coast' · 'Italy' · 'Israel' · 'Ireland' · 'Iraq' · 'Iran' · 'Indonesia' · 'India' · 'Iceland' · 'Hungary' · 'Honduras' · 'Haiti' · 'Guyana' · 'Guinea-Bissau' · 'Guinea' · 'Guatemala' · 'Grenada' · 'Greenland' · 'Greece' · 'Ghana' · 'Germany' · 'Georgia[h]' · 'Gambia' · 'Gabon' · 'France[f][g]' · 'Finland' · 'Fiji' · 'Faroe Islands' · 'Ethiopia' · 'Eswatini' · 'Estonia' · 'Equatorial Guinea' · 'El Salvador' · 'Egypt' · 'Ecuador' · 'DR Congo' · 'Dominican Republic' · 'Dominica' · 'Djibouti' · 'Denmark[e]' · 'Czechia' · 'Cyprus[d]' · 'Cuba' · 'Croatia' · 'Costa Rica' · 'Colombia' · 'China[c]' · 'Chile' · 'Chad' · 'Canada' · 'Cameroon' · 'Cambodia' · 'Burundi' · 'Burkina Faso' · 'Bulgaria' · 'Brunei' · 'Brazil' · 'Botswana' · 'Bosnia and Herzegovina' · 'Bolivia' · 'Bhutan' · 'Benin' · 'Belize' · 'Belgium' · 'Belarus' · 'Barbados' · 'Bangladesh' · 'Bahrain' · 'Bahamas' · 'Azerbaijan' · 'Austria' · 'Australia' · 'Armenia' · 'Argentina' · 'Antigua and Barbuda' · 'Angola' · 'Andorra' · 'Algeria' · 'Albania' · 'Afghanistan'

TASK 7: Identify countries names with a specific pattern

The goal of task 7 is using a regular expression to find any countires start with `United`

```
# Find matches using regular expression
matches <- grepl("United.+", table1$country)

# Print the matched country names
matched_countries <- table1$country[matches]
print(matched_countries)
```

```
[1] "United Arab Emirates" "United Kingdom"       "United States"
```

TASK 8: Pick two countries you are interested, and then review their testing data

The goal of task 8 is to compare the COVID-19 test data between two countires, you will need to select two rows from the dataframe, and select country, confirmed, confirmed-population-ratio columns

```
# Select a country name
selected_country <- "United States"

# Select columns for the subset
selected_columns <- c("country", "date", "confirmed", "confirmed.populati

# Create a subset based on the selected country and columns
subset_df <- table1[table1$country == selected_country, selected_columns]

subset_df
```

A tibble: 1 × 4

country	date	confirmed	confirmed.population.ratio
<chr>	<chr>	<dbl>	<chr>
United States	29 Jul 2022	90749469	27.4

```
# Select a country name
selected_country <- "United Kingdom"

# Select columns for the subset
selected_columns <- c("country", "date", "confirmed", "confirmed.popula

# Create a subset based on the selected country and columns
subset_df <- table1[table1$country == selected_country, selected_columns

subset_df
```

A tibble: 1 × 4

country	date	confirmed	confirmed.population.ratio
<chr>	<chr>	<dbl>	<chr>
United Kingdom	19 May 2022	22232377	32.9

TASK 9: Compare which one of the selected countries has a larger ratio of confirmed cases to population

The goal of task 9 is to find out which country you have selected before has larger ratio of confirmed cases to population, which may indicate that country has higher COVID-19 infection risk

```
# Read the data from the CSV file into a data frame
data <- read.csv("table1_data.csv")

# Selected countries
selected_country1 <- "United Kingdom"
selected_country2 <- "United States"

# Filter the data for the selected countries
filtered_data <- data[data$country %in% c(selected_country1, selected_c

# Calculate the ratio of confirmed cases to population for each country
country1_ratio <- filtered_data$confirmed.population.ratio[filtered_data
country2_ratio <- filtered_data$confirmed.population.ratio[filtered_data

# Compare the ratios and print the result
if (country1_ratio > country2_ratio) {
  cat(paste0(selected_country1, " has a larger ratio of confirmed cases
} else if (country2_ratio > country1_ratio) {
  cat(paste0(selected_country2, " has a larger ratio of confirmed cases
} else {
  cat("Both countries have the same ratio of confirmed cases to populat
}
```

```
United Kingdom has a larger ratio of confirmed cases to population.
```

```
colnames(table1)
```

'country' · 'date' · 'tested' · 'confirmed' · 'confirmed.tested ratio' · 'tested.population.ratio' · 'confirmed.population.ratio'

TASK 10: Find countries with confirmed to population ratio rate less than a threshold

The goal of task 10 is to find out which countries have the confirmed to population ratio less than 1%, it may indicate the risk of those countries are relatively low

```
# Specify the threshold for confirmed to population ratio
threshold <- 1

# Create a subset based on the threshold
subset_df <- table1[table1$`confirmed.population.ratio` < threshold, "cou

# Print the countries with the confirmed to population ratio less than th
print(subset_df)
```

```
# A tibble: 53 × 1
   country
   <chr>
 1 Afghanistan
 2 Algeria
 3 Angola
 4 Antigua and Barbuda
 5 Bangladesh
 6 Benin
 7 Brunei
 8 Burkina Faso
 9 Burundi
10 Cambodia
# ℹ 43 more rows
```