Name: Ong Jing Wei

Student ID: 32909764

Generative AI was used in this assignment.

1. Descriptive analysis and pre-processing. (6 Marks)

(a) Describe the data overall, including things such as dimension, data types, distribution of numerical attributes, variety of non-numerical (text) attributes, missing values, and anything else of interest or relevance.

This dataset consists of **40000** rows and **52** columns.

There are 52 attributes in this dataset where **45** attributes which belongs to the **integer** class are:

"**employstatus_1**", "**employstatus_2**", "**employstatus_3**", "**employstatus_4**", "**employstatus_5**", "**employstatus_6**", "**employstatus_7**", "**employstatus_8**", "**employstatus_9**", "**employstatus_10**", "**isoFriends_inPerson**", "**isoOthPpl_inPerson**", "**isoFriends_online**", "**isoOthPpl_online**", "**lone01**", "**lone02**", "**lone03**", "**happy**", "**lifeSat**", "**MLQ**", "**bor01**", "**bor02**", "**bor03**", "**consp01**", "**consp02**", "**consp03**", "**c19perBeh01**", "**c19perBeh02**", "**c19perBeh03**", "**c19RCA01**", "**c19RCA02**", "**c19RCA03**", "**coronaClose_1**", "**coronaClose_2**", "**coronaClose_3**", "**coronaClose_4**", "**coronaClose_5**", "**coronaClose_6**", "**gender**", "**age**", "**edu**", "**c19ProSo01**", "**c19ProSo02**", "**c19ProSo03**", "**c19ProSo04**"

and **7** attributes which belongs to the **character** class are:

"**rankOrdLife_1**", "**rankOrdLife_2**", "**rankOrdLife_3**", "**rankOrdLife_4**", "**rankOrdLife_5**", "**rankOrdLife_6**", "**coded_country**"

```
> summary(covid)
 employstatus_1  employstatus_2  employstatus_3  employstatus_4  employstatus_5  employstatus_6  employstatus_7  employstatus_8  employstatus_9  employstatus_10
 Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1
 1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1
 Median :1       Median :1       Median :1       Median :1       Median :1       Median :1       Median :1       Median :1       Median :1       Median :1
 Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1
 3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1
 Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1
 NA's   :34362   NA's   :33277   NA's   :29151   NA's   :36452   NA's   :37972   NA's   :36892   NA's   :36439   NA's   :39272   NA's   :31782   NA's   :39073
 isoFriends_inPerson isoOthPpl_inPerson isoFriends_online isoOthPpl_online     lone01          lone02          lone03          happy           lifeSat
 Min.   :0.000       Min.   :0.00       Min.   :0.000     Min.   :0.000    Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   : 1.000   Min.   :1.000
 1st Qu.:0.000       1st Qu.:0.00       1st Qu.:2.000     1st Qu.:0.000    1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.: 5.000   1st Qu.:3.000
 Median :1.000       Median :1.00       Median :5.000     Median :2.000    Median :2.000   Median :3.000   Median :2.000   Median : 7.000   Median :4.000
 Mean   :2.073       Mean   :1.96       Mean   :4.397     Mean   :2.866    Mean   :2.418   Mean   :2.664   Mean   :2.078   Mean   : 6.332   Mean   :4.136
 3rd Qu.:4.000       3rd Qu.:3.00       3rd Qu.:7.000     3rd Qu.:5.000    3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.: 8.000   3rd Qu.:5.000
 Max.   :7.000       Max.   :7.00       Max.   :7.000     Max.   :7.000    Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :10.000   Max.   :6.000
 NA's   :320         NA's   :520        NA's   :942       NA's   :1154     NA's   :77      NA's   :119     NA's   :137     NA's   :500      NA's   :107
      MLQ             bor01           bor02           bor03           consp01         consp02         consp03       rankOrdLife_1      rankOrdLife_2
 Min.   :-3.000   Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.0000   Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Length:40000       Length:40000
 1st Qu.: 0.000   1st Qu.:-1.0000   1st Qu.:-2.0000   1st Qu.:-1.0000   1st Qu.: 5.00   1st Qu.: 5.000   1st Qu.: 4.000   Class :character   Class :character
 Median : 1.000   Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 7.00   Median : 8.000   Median : 5.000   Mode  :character   Mode  :character
 Mean   : 0.842   Mean   : 0.3222   Mean   : 0.0421   Mean   : 0.3141   Mean   : 6.83   Mean   : 7.135   Mean   : 5.577
 3rd Qu.: 2.000   3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 9.00   3rd Qu.: 9.000   3rd Qu.: 8.000
 Max.   : 3.000   Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.0000   Max.   :10.00   Max.   :10.000   Max.   :10.000
 NA's   :109      NA's   :144       NA's   :164       NA's   :167       NA's   :1519    NA's   :1544     NA's   :1565

 rankOrdLife_3      rankOrdLife_4      rankOrdLife_5      rankOrdLife_6      c19perBeh01      c19perBeh02      c19perBeh03      c19RCA01         c19RCA02
 Length:40000       Length:40000       Length:40000       Length:40000       Min.   :-3.000   Min.   :-3.000   Min.   :-3.00   Min.   :-3.000   Min.   :-3.000
 Class :character   Class :character   Class :character   Class :character   1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.: 1.00   1st Qu.: 0.000   1st Qu.: 2.000
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median : 3.000   Median : 3.000   Median : 2.00   Median : 2.000   Median : 3.000
                                                                            Mean   : 2.313   Mean   : 2.424   Mean   : 1.83   Mean   : 1.243   Mean   : 2.044
                                                                            3rd Qu.: 3.000   3rd Qu.: 3.000   3rd Qu.: 3.00   3rd Qu.: 3.000   3rd Qu.: 3.000
                                                                            Max.   : 3.000   Max.   : 3.000   Max.   : 3.00   Max.   : 3.000   Max.   : 3.000
                                                                            NA's   :122      NA's   :126      NA's   :128     NA's   :121      NA's   :128
    c19RCA03      coronaClose_1   coronaClose_2   coronaClose_3   coronaClose_4   coronaClose_5   coronaClose_6      gender           age             edu
 Min.   :-3.000   Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1       Min.   :1.00   Min.   :1.000   Min.   :1.000
 1st Qu.: 0.000   1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1       1st Qu.:1.00   1st Qu.:2.000   1st Qu.:4.000
 Median : 2.000   Median :1       Median :1       Median :1       Median :1       Median :1       Median :1       Median :1.00   Median :3.000   Median :5.000
 Mean   : 1.154   Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1       Mean   :1.39   Mean   :2.893   Mean   :4.394
 3rd Qu.: 3.000   3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:1       3rd Qu.:2.00   3rd Qu.:4.000   3rd Qu.:5.000
 Max.   : 3.000   Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :1       Max.   :3.00   Max.   :8.000   Max.   :7.000
 NA's   :137      NA's   :39447   NA's   :38757   NA's   :38443   NA's   :35096   NA's   :35497   NA's   :10657   NA's   :225    NA's   :248     NA's   :270
```

coded_country	c19ProSo01	c19ProSo02	c19ProSo03	c19ProSo04
Length:40000	Min. :-3.0000	Min. :-3.0000	Min. :-3.0000	Min. :-3.000
Class :character	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.000
Mode :character	Median : 1.0000	Median : 1.0000	Median : 1.0000	Median : 2.000
	Mean : 0.9692	Mean : 0.6672	Mean : 0.5451	Mean : 1.274
	3rd Qu.: 2.0000	3rd Qu.: 2.0000	3rd Qu.: 2.0000	3rd Qu.: 2.000
	Max. : 3.0000	Max. : 3.0000	Max. : 3.0000	Max. : 3.000
	NA's :114	NA's :129	NA's :144	NA's :142

By looking from the summary and the boxplot (Figure1.1.1, Figure1.1.2, Figure1.1.3, Figure1.1.4), we can see that the employment status attributes from **employstatus_1** to **employstatus_10** are binary attributes as the minimum and maximum values of 1 across these attributes suggest a binary coding scheme where 1 indicates the label of that the variable best describes the individual's employment status. The presence of missing values implies that respondents were likely instructed to select only one option among the available employment status categories, while not applicable is set to a null value. This is same with the attributes **coronaClose_1** to **coronaClose_6** as the minimum and maximum value of these variables are also 1, the presence of missing values implies that respondents were likely instructed to select only one option from the available categories, leaving other options unselected. Hence the missing values here are reasonable.

The attributes **isoFriends_inPerson**, **isoOthPpl_inPerson**, **isoFriends_online**, **isoOthPpl_online, lone01** to **lone03**, **happy**, **lifeSat**, **consp01** to **consp03** are attributes with positive range values, where higher values indicate greater agreement or preference towards the category. While the variables **MLQ**, **bor01** to **bor03**, **c19perBeh01** to **c19perBeh03**, **c19RCA01** to **c19RCA03**, **c19ProSo01** to **c19ProSo04** are attributes with the value ranging from negative to positive, where the values indicating the extent of agreement or disagreement towards the category.

isoFriends_inPerson, **isoOthPpl_inPerson**, **isoFriends_online**, **isoOthPpl_online** have a high number of missing values, suggesting that the respondent might not want to disclose personal information about their social interactions. Similarly, the high number of missing values in the **happy** attribute might be due to the impact of pandemic, which could make it difficult for individuals to accurately assess their happiness. Additionally, for the high number of missing values in attributes **consp01** to **consp03** may indicate concerns about potential repercussions, such as being subject to surveillance or facing criticism for expressing distrust in political leaders or institutions.

rankOrdLife_1 to **rankOrdLife_6** is the categorical attributes where the values are encoded using letters A to F, each corresponding to a specific aspect of life. Higher-ranked variable indicates greater importance or significance in the respondent's lives. These variables can be used to explore correlations between these rankings and other variables.

gender, age and **edu** are also categorical attributes where it uses integer to represent distinct categories or groups. Missing values in these attributes suggest that respondents decided not to disclose their preferences or attitudes for the corresponding categories.

	Var1	Freq
105	United States of America	7007
94	Spain	1996
72	Netherlands	1922
40	Greece	1837
86	Romania	1683
46	Indonesia	1542
85	Republic of Serbia	1330
51	Italy	1254
104	United Kingdom	1214
101	Turkey	1107
37	France	1085
39	Germany	1052
23	China	1006
21	Canada	970
81	Philippines	963
87	Russia	921
88	Saudi Arabia	908
93	South Korea	897
92	South Africa	896
102	Ukraine	896
5	Argentina	868
17	Brazil	864
53	Japan	835

7	Australia	768
32	Egypt	710
62	Malaysia	547
56	Kosovo	511
55	Kazakhstan	487
82	Poland	472
77	Pakistan	470
43	Hungary	265
22	Chile	218
26	Croatia	206
47	Iran	206
80	Peru	193
42	Hong Kong S.A.R.	186
1		157
109	Vietnam	147
89	Singapore	141
3	Algeria	117
11	Bangladesh	99
98	Thailand	97
97	Taiwan	94
45	India	60
103	United Arab Emirates	57
27	Cyprus	51
95	Sweden	49
50	Israel	48
13	Belgium	42

96	Switzerland	41
100	Tunisia	40
83	Portugal	34
8	Austria	33
33	El Salvador	30
66	Mexico	28
24	Colombia	26
70	Morocco	23
48	Iraq	20
67	Moldova	19
49	Ireland	18
99	Trinidad and Tobago	18
73	New Zealand	17
61	Luxembourg	13
29	Denmark	12
78	Palestine	12
28	Czech Republic	11
36	Finland	11
63	Mali	11
15	Bosnia and Herzegovina	9
19	Bulgaria	9
60	Lithuania	8
75	Norway	8
108	Venezuela	8
90	Slovakia	7
69	Montenegro	6

10	Bahrain	5
38	Georgia	5
52	Jamaica	5
54	Jordan	5
106	Uruguay	5
31	Ecuador	4
59	Lebanon	4
2	Albania	3
18	Brunei	3
25	Costa Rica	3
30	Dominican Republic	3
74	Nigeria	3
12	Belarus	2
34	Estonia	2
41	Guatemala	2
57	Kuwait	2
79	Panama	2
84	Qatar	2
107	Uzbekistan	2
4	Andorra	1
6	Armenia	1
9	Azerbaijan	1
14	Benin	1
16	Botswana	1
20	Cameroon	1
35	Ethiopia	1

44	Iceland	1
58	Kyrgyzstan	1
64	Malta	1
65	Mauritius	1
68	Mongolia	1
71	Myanmar	1
76	Oman	1
91	Slovenia	1

The **coded_country** attribute consists of a list of countries with a total of 109 distinct countries. Notably, most of the respondents are from the United States of America, with a frequency count of 7007. While Andorra, Armenia, Azerbaijan, Benin, Botswana, Cameroon, Ethiopia, Iceland, Kyrgyzstan, Malta, Mauritius, Mongolia, Myanmar, Oman, and Slovenia have the least frequency which is only 1. However, there is an empty value observed between Hong Kong S.A.R and Vietnam, occurring 157 times. This suggests that there might be some incomplete values for this attribute, or the respondent might not want to disclose where they are from.

(b) Comment on any pre-processing or data manipulation required for the following analysis.

For data pre-processing, I addressed missing values in several attributes. Initially, I replaced NULL values in **employstatus_1** to **employstatus_10** and **coronaClose_1** to **coronaClose_6** with 0. This adjustment was necessary as respondents were instructed to select only one option from available categories, leaving others unselected. To avoid dropping entire rows due to these attributes, I replaced NA values accordingly. Additionally, I converted empty values in the **coded_country** attribute to NA. Since empty values do not contribute to our analysis, replacing them with NA facilitated their removal in subsequent steps. After handling missing values, I performed data clean up by dropping rows with NA values, preparing the dataset for further analysis.

Furthermore, I encoded categorical attributes **rankOrdLife_1** to **rankOrdLife_6** into numeric values for easier comparison. This involved defining a function to encode categorical variables and applying it to each column using a mapping scheme.

2. Focus country vs all other countries as a group. (12 Marks)

(a) Identify your focus country from the accompanying list (FocusCountryByID.pdf). How do participant responses (attributes) for your focus country differ from the other countries in the survey as a group?

When comparing the summary statistics below and the boxplot (Figure 2.1.1, Figure 2.1.2, Figure 2.1.3) between the Netherlands and other countries, we observe distinct patterns. In the Netherlands, a significant proportion of respondents are employed, and working 24-39 hours per week with around 75% as indicated by the 3rd quartile of employstatus_2. The mean of 0.3001 suggests that approximately 30% of respondents fall into this category, reflecting a common employment pattern. Conversely, in other countries, respondents are predominantly working 40 or more hours per week.

Moreover, respondents from the Netherlands appear to have more social interactions with individuals outside their households compared to those from other countries. This is evident from higher mean, median, and 3rd quartile values in isoOthPpl_inPerson. Additionally, the p-value for the t-test below is extremely small also indicates that there is significantly larger mean value of isoOthPpl_inPerson in the Netherlands compared to other countries, further highlighting this distinction.

```
> # t-test for isoOthPpl_inPerson
> t_test_isoOthPpl_inPerson <- t.test(netherlands_data$isoOthPpl_inPerson, other_countries_data$isoOthPpl_inP
erson)
> t_test_isoOthPpl_inPerson

        Welch Two Sample t-test

data:  netherlands_data$isoOthPpl_inPerson and other_countries_data$isoOthPpl_inPerson
t = 10.883, df = 1538.9, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.5575795 0.8027546
sample estimates:
mean of x mean of y
 2.525471  1.845304
```

Respondents from Netherlands show lower tendencies towards feelings of isolation and exclusion, as indicated by lower median and mean values in lone02 and lone03 compared to respondents from other countries. Similarly, the mean values for conspiracy beliefs (consp01, consp02, consp03) in the Netherlands are lower than in other countries, suggesting a lower inclination towards believing in conspiracies. This implies that individuals in the Netherlands may have higher levels of trust in public information, political transparency, and government agencies' actions compared to respondents from other countries. A t-test conducted for one of the conspiracy attributes (consp02) reveals an extremely small p-value, indicating a significant difference in mean values between the Netherlands and other countries.

```
> # t-test for consp02
> t_test_consp02 <- t.test(netherlands_data$consp02, other_countries_data$consp02)
> t_test_consp02

        Welch Two Sample t-test

data:  netherlands_data$consp02 and other_countries_data$consp02
t = -24.907, df = 1546, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.846184 -1.576630
sample estimates:
mean of x mean of y
 5.625262  7.336668
```

```
> summary(netherlands_data)
 employstatus_1   employstatus_2   employstatus_3   employstatus_4    employstatus_5    employstatus_6
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
 Mean   :0.1445   Mean   :0.3001   Mean   :0.1961   Mean   :0.05024   Mean   :0.02442   Mean   :0.06909
 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
 employstatus_7   employstatus_8    employstatus_9   employstatus_10   isoFriends_inPerson isoOthPpl_inPerson
 Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.000       Min.   :0.000
 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.000       1st Qu.:0.000
 Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000   Median :1.000       Median :2.000
 Mean   :0.1117   Mean   :0.05652   Mean   :0.1689   Mean   :0.04676   Mean   :1.909       Mean   :2.525
 3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:3.000       3rd Qu.:4.000
 Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000   Max.   :7.000       Max.   :7.000
 isoFriends_online isoOthPpl_online     lone01          lone02          lone03         happy
 Min.   :0.00      Min.   :0.000    Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   : 1.000
 1st Qu.:2.00      1st Qu.:0.000    1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.00   1st Qu.: 6.000
 Median :4.00      Median :2.000    Median :2.000   Median :2.000   Median :1.00   Median : 7.000
 Mean   :3.88      Mean   :2.522    Mean   :1.999   Mean   :2.397   Mean   :1.78   Mean   : 6.871
 3rd Qu.:7.00      3rd Qu.:5.000    3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.00   3rd Qu.: 8.000
 Max.   :7.00      Max.   :7.000    Max.   :5.000   Max.   :5.000   Max.   :5.00   Max.   :10.000
    lifeSat           MLQ              bor01           bor02           bor03            consp01
 Min.   :1.000   Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.0000   Min.   : 0.000
 1st Qu.:4.000   1st Qu.: 0.0000   1st Qu.:-2.0000   1st Qu.:-2.0000   1st Qu.:-1.0000   1st Qu.: 4.000
 Median :5.000   Median : 1.0000   Median : 0.0000   Median : 0.0000   Median : 1.0000   Median : 6.000
 Mean   :4.574   Mean   : 0.8255   Mean   :-0.0321   Mean   :-0.3043   Mean   : 0.4955   Mean   : 5.846
 3rd Qu.:5.000   3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 1.0000   3rd Qu.: 2.0000   3rd Qu.: 8.000
 Max.   :6.000   Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.0000   Max.   :10.000
    consp02          consp03       rankOrdLife_1   rankOrdLife_2   rankOrdLife_3   rankOrdLife_4
 Min.   : 0.000   Min.   : 0.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.:4.000   1st Qu.:3.000   1st Qu.:4.000   1st Qu.:2.000
 Median : 6.000   Median : 5.000   Median :5.000   Median :4.000   Median :5.000   Median :2.000
 Mean   : 5.625   Mean   : 4.418   Mean   :4.385   Mean   :3.999   Mean   :4.861   Mean   :2.436
 3rd Qu.: 8.000   3rd Qu.: 6.000   3rd Qu.:6.000   3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:3.000
 Max.   :10.000   Max.   :10.000   Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000

rankOrdLife_5   rankOrdLife_6    c19perBeh01      c19perBeh02      c19perBeh03        c19RCA01
Min.   :1.000   Min.   :1.000   Min.   :-3.000   Min.   :-3.000   Min.   :-3.000   Min.   :-3.000
1st Qu.:1.000   1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 0.000
Median :2.000   Median :3.000   Median : 2.000   Median : 3.000   Median : 2.000   Median : 2.000
Mean   :2.149   Mean   :3.171   Mean   : 2.112   Mean   : 2.373   Mean   : 1.278   Mean   : 0.993
3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.: 3.000   3rd Qu.: 3.000   3rd Qu.: 2.000   3rd Qu.: 3.000
Max.   :6.000   Max.   :6.000   Max.   : 3.000   Max.   : 3.000   Max.   : 3.000   Max.   : 3.000
   c19RCA02         c19RCA03       coronaClose_1    coronaClose_2    coronaClose_3    coronaClose_4
Min.   :-3.000   Min.   :-3.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
1st Qu.: 1.000   1st Qu.:-2.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
Median : 2.000   Median : 0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.000
Mean   : 1.683   Mean   : 0.1047   Mean   :0.0307   Mean   :0.0642   Mean   :0.0635   Mean   :0.164
3rd Qu.: 3.000   3rd Qu.: 2.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000
Max.   : 3.000   Max.   : 3.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
coronaClose_5    coronaClose_6        gender           age              edu       coded_country
Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :1.000   Min.   :1.00   Length:1433
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:3.00   Class :character
Median :0.0000   Median :1.0000   Median :1.000   Median :3.000   Median :4.00   Mode  :character
Mean   :0.1486   Mean   :0.6315   Mean   :1.369   Mean   :3.338   Mean   :4.39
3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:5.000   3rd Qu.:6.00
Max.   :1.0000   Max.   :1.0000   Max.   :3.000   Max.   :8.000   Max.   :7.00
  c19ProSo01       c19ProSo02       c19ProSo03       c19ProSo04
Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.000
1st Qu.: 0.0000   1st Qu.:-1.0000   1st Qu.: 0.0000   1st Qu.: 0.000
Median : 1.0000   Median : 0.0000   Median : 1.0000   Median : 2.000
Mean   : 0.9525   Mean   : 0.2114   Mean   : 0.6364   Mean   : 1.337
3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 2.000
Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.000
```

```
> summary(other_countries_data)
 employstatus_1   employstatus_2   employstatus_3   employstatus_4    employstatus_5   employstatus_6
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000
 Mean   :0.1424   Mean   :0.1608   Mean   :0.2767   Mean   :0.09127   Mean   :0.0528   Mean   :0.07864
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
 employstatus_7  employstatus_8   employstatus_9   employstatus_10   isoFriends_inPerson
 Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.000
 1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.000
 Median :0.000   Median :0.0000   Median :0.0000   Median :0.00000   Median :1.000
 Mean   :0.091   Mean   :0.0162   Mean   :0.2103   Mean   :0.02272   Mean   :1.966
 3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:3.000
 Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :7.000
 isoOthPpl_inPerson isoFriends_online isoOthPpl_online     lone01          lone02        lone03
 Min.   :0.000      Min.   :0.000     Min.   :0.000     Min.   :1.000   Min.   :1.0   Min.   :1.00
 1st Qu.:0.000      1st Qu.:2.000     1st Qu.:0.000     1st Qu.:1.000   1st Qu.:2.0   1st Qu.:1.00
 Median :1.000      Median :5.000     Median :2.000     Median :2.000   Median :3.0   Median :2.00
 Mean   :1.845      Mean   :4.454     Mean   :2.867     Mean   :2.436   Mean   :2.7   Mean   :2.08
 3rd Qu.:3.000      3rd Qu.:7.000     3rd Qu.:5.000     3rd Qu.:3.000   3rd Qu.:4.0   3rd Qu.:3.00
 Max.   :7.000      Max.   :7.000     Max.   :7.000     Max.   :5.000   Max.   :5.0   Max.   :5.00
     happy           lifeSat           MLQ             bor01             bor02
 Min.   : 1.000   Min.   :1.000   Min.   :-3.000   Min.   :-3.0000   Min.   :-3.00000
 1st Qu.: 5.000   1st Qu.:3.000   1st Qu.: 0.000   1st Qu.:-1.0000   1st Qu.:-2.00000
 Median : 7.000   Median :4.000   Median : 1.000   Median : 0.0000   Median : 0.00000
 Mean   : 6.331   Mean   :4.145   Mean   : 0.867   Mean   : 0.3217   Mean   : 0.05991
 3rd Qu.: 8.000   3rd Qu.:5.000   3rd Qu.: 2.000   3rd Qu.: 2.0000   3rd Qu.: 2.00000
 Max.   :10.000   Max.   :6.000   Max.   : 3.000   Max.   : 3.0000   Max.   : 3.00000
     bor03            consp01          consp02          consp03        rankOrdLife_1   rankOrdLife_2
 Min.   :-3.0000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   :1.000   Min.   :1.000
 1st Qu.:-1.0000   1st Qu.: 5.000   1st Qu.: 6.000   1st Qu.: 4.000   1st Qu.:4.000   1st Qu.:3.000
 Median : 0.0000   Median : 7.000   Median : 8.000   Median : 6.000   Median :5.000   Median :4.000
 Mean   : 0.2955   Mean   : 6.992   Mean   : 7.337   Mean   : 5.691   Mean   :4.659   Mean   :3.531
 3rd Qu.: 2.0000   3rd Qu.: 9.000   3rd Qu.:10.000   3rd Qu.: 8.000   3rd Qu.:6.000   3rd Qu.:5.000
 Max.   : 3.0000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :6.000   Max.   :6.000


 rankOrdLife_3   rankOrdLife_4   rankOrdLife_5   rankOrdLife_6    c19perBeh01      c19perBeh02
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :-3.000   Min.   :-3.000
 1st Qu.:4.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 2.000
 Median :5.000   Median :3.000   Median :2.000   Median :3.000   Median : 3.000   Median : 3.000
 Mean   :4.748   Mean   :2.725   Mean   :2.098   Mean   :3.239   Mean   : 2.372   Mean   : 2.483
 3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.: 3.000   3rd Qu.: 3.000
 Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   : 3.000   Max.   : 3.000
  c19perBeh03        c19RCA01         c19RCA02         c19RCA03      coronaClose_1     coronaClose_2
 Min.   :-3.000   Min.   :-3.000   Min.   :-3.000   Min.   :-3.000   Min.   :0.00000   Min.   :0.00000
 1st Qu.: 1.000   1st Qu.: 0.000   1st Qu.: 2.000   1st Qu.: 0.000   1st Qu.:0.00000   1st Qu.:0.00000
 Median : 2.000   Median : 2.000   Median : 3.000   Median : 2.000   Median :0.00000   Median :0.00000
 Mean   : 1.894   Mean   : 1.296   Mean   : 2.122   Mean   : 1.188   Mean   :0.01033   Mean   :0.02862
 3rd Qu.: 3.000   3rd Qu.: 3.000   3rd Qu.: 3.000   3rd Qu.: 3.000   3rd Qu.:0.00000   3rd Qu.:0.00000
 Max.   : 3.000   Max.   : 3.000   Max.   : 3.000   Max.   : 3.000   Max.   :1.00000   Max.   :1.00000
 coronaClose_3     coronaClose_4    coronaClose_5    coronaClose_6        gender           age
 Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :1.000
 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:2.000
 Median :0.00000   Median :0.0000   Median :0.0000   Median :1.0000   Median :1.000   Median :3.000
 Mean   :0.03642   Mean   :0.1236   Mean   :0.1119   Mean   :0.7411   Mean   :1.386   Mean   :2.901
 3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:4.000
 Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :3.000   Max.   :8.000
      edu        coded_country        c19ProSo01       c19ProSo02       c19ProSo03
 Min.   :1.000   Length:33582       Min.   :-3.0000   Min.   :-3.0000   Min.   :-3.000
 1st Qu.:4.000   Class :character   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.000
 Median :5.000   Mode  :character   Median : 1.0000   Median : 1.0000   Median : 1.000
 Mean   :4.404                      Mean   : 0.9845   Mean   : 0.6743   Mean   : 0.541
 3rd Qu.:5.000                      3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 2.000
 Max.   :7.000                      Max.   : 3.0000   Max.   : 3.0000   Max.   : 3.000
   c19ProSo04
 Min.   :-3.000
 1st Qu.: 1.000
 Median : 2.000
 Mean   : 1.339
 3rd Qu.: 3.000
 Max.   : 3.000
```

(b) How well do participant responses (attributes) predict pro-social attitudes **(c19ProSo01,2,3 and 4)** for your focus country? Which attributes seem to be the best predictors? Explain your reasoning.

c19ProSo01		
1st Best Predictor	c19perBeh02, p-value: 0.0000621	
2nd Best Predictor	isoOthPpl_inPerson, p-value: 0.000678	
R-squared: 0.3381	RSE: 1.2	
c19ProSo02		
1st Best Predictor	c19RCA01, p-value: 0.000000579	
2nd Best Predictor	employstatus_7, p-value: 0.0000903	
R-squared: 0.2912	RSE: 1.443	
c19ProSo03		
1st Best Predictor	age, p-value: 0.000000344	
2nd Best Predictor	Employstatus_9, p-value: 0.02640	
R-squared: 0.4557	RSE: 1.306	
c19ProSo04		
1st Best Predictor	c19perBeh03, p-value: 0.000000418	
2nd Best Predictor	employstatus_9, p-value: 0.000342	
R-squared: 0.442	RSE: 1.17	

The table above summarizes the best predictors for all pro-social attitudes based on the linear regression analysis. These predictors were chosen because of their significantly low p-values, suggesting a strong correlation with the response variable. This indicates compelling evidence of their influence on the respondents' pro-social attitudes.

The average R-squared value, around 0.38, suggests that the model accounts for a moderate portion of the variability in the response variable. While this indicates some level of usefulness, it also implies that the model may not fully capture all underlying relationships in the data.

Furthermore, the low average residual standard error, approximately 1.28, implies that the model fits the data well and can make accurate predictions. This suggests that the model's predictions are generally close to the actual values, indicating its reliability in predicting pro-social attitudes based on the chosen predictors.

```
Call:
lm(formula = c19ProSo01 ~ ., data = netherlands_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1439 -0.6452  0.1016  0.7854  3.5074

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.3739872  0.5480921  -0.682 0.495136
employstatus_1     -0.1412853  0.1201039  -1.176 0.239654
employstatus_2     -0.0258750  0.1272813  -0.203 0.838938
employstatus_3     -0.0244965  0.1406741  -0.174 0.861783
employstatus_4     -0.0304483  0.1677422  -0.182 0.855987
employstatus_5      0.0927445  0.2189730   0.424 0.671965
employstatus_6     -0.1447712  0.1425236  -1.016 0.309916
employstatus_7     -0.3633045  0.1623902  -2.237 0.025430 *
employstatus_8     -0.3035025  0.1720180  -1.764 0.077891 .
employstatus_9      0.2416462  0.1335435   1.809 0.070591 .
employstatus_10     0.1607027  0.1572076   1.022 0.306850
isoFriends_inPerson 0.0218011  0.0169090   1.289 0.197502
isoOthPpl_inPerson  0.0543594  0.0159595   3.406 0.000678 ***
isoFriends_online   0.0128673  0.0147736   0.871 0.383924
isoOthPpl_online   -0.0167216  0.0143444  -1.166 0.243929
lone01              0.0924839  0.0478659   1.932 0.053545 .
lone02              0.0147952  0.0406912   0.364 0.716215
lone03             -0.0656983  0.0440279  -1.492 0.135875
happy              -0.0400213  0.0318541  -1.256 0.209186
lifeSat             0.0304800  0.0505524   0.603 0.546648
MLQ                -0.0106343  0.0269340  -0.395 0.693031
bor01               0.0020169  0.0250294   0.081 0.935786
bor02              -0.0094411  0.0259699  -0.364 0.716257
bor03               0.0134239  0.0214530   0.626 0.531591
consp01             0.0142699  0.0176812   0.807 0.419767
consp02             0.0233618  0.0191450   1.220 0.222576
consp03            -0.0041894  0.0156625  -0.267 0.789141
rankOrdLife_1      -0.0276882  0.0296229  -0.935 0.350112
rankOrdLife_2       0.0006791  0.0326239   0.021 0.983397
rankOrdLife_3      -0.0091495  0.0336959  -0.272 0.786023
rankOrdLife_4      -0.0094076  0.0378821  -0.248 0.803910
rankOrdLife_5      -0.0298346  0.0329224  -0.906 0.364984
rankOrdLife_6              NA         NA      NA       NA
c19perBeh01         0.0392345  0.0343554   1.142 0.253644
c19perBeh02         0.1793836  0.0446541   4.017 6.21e-05 ***
c19perBeh03        -0.0742231  0.0251008  -2.957 0.003159 **
c19RCA01            0.0165960  0.0201593   0.823 0.410512
c19RCA02            0.0055002  0.0292648   0.188 0.850948
c19RCA03           -0.0399355  0.0209970  -1.902 0.057383 .
coronaClose_1       0.3801891  0.2226826   1.707 0.087988 .
coronaClose_2      -0.0882258  0.1616685  -0.546 0.585346
coronaClose_3      -0.0342230  0.1539486  -0.222 0.824112
coronaClose_4       0.2792447  0.1311729   2.129 0.033445 *
coronaClose_5       0.1174055  0.1357323   0.865 0.387201
coronaClose_6       0.0254124  0.1424683   0.178 0.858456
gender              0.0991188  0.0685773   1.445 0.148584
age                 0.0562947  0.0321358   1.752 0.080034 .
edu                 0.0270482  0.0253371   1.068 0.285918
c19ProSo02          0.1261517  0.0221105   5.706 1.42e-08 ***
c19ProSo03          0.3341830  0.0230192  14.518  < 2e-16 ***
c19ProSo04          0.0575319  0.0275390   2.089 0.036881 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.2 on 1383 degrees of freedom
Multiple R-squared:  0.3381,    Adjusted R-squared:  0.3147
F-statistic: 14.42 on 49 and 1383 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo02 ~ ., data = netherlands_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6165 -0.9706  0.1850  1.0320  3.9915

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -3.691797   0.651449  -5.667 1.77e-08 ***
employstatus_1      0.293672   0.144232   2.036 0.041930 *
employstatus_2      0.375359   0.152673   2.459 0.014071 *
employstatus_3      0.500717   0.168568   2.970 0.003025 **
employstatus_4      0.169070   0.201593   0.839 0.401800
employstatus_5      0.022567   0.263242   0.086 0.931696
employstatus_6      0.167909   0.171331   0.980 0.327245
employstatus_7      0.763722   0.194480   3.927 9.03e-05 ***
employstatus_8      0.040813   0.207011   0.197 0.843737
employstatus_9     -0.017983   0.160721  -0.112 0.910926
employstatus_10    -0.122325   0.189021  -0.647 0.517641
isoFriends_inPerson 0.028805   0.020324   1.417 0.156612
isoOthPpl_inPerson -0.019017   0.019258  -0.987 0.323590
isoFriends_online   0.016121   0.017759   0.908 0.364156
isoOthPpl_online    0.038127   0.017221   2.214 0.026994 *
lone01              0.098043   0.057557   1.703 0.088713 .
lone02             -0.092609   0.048854  -1.896 0.058216 .
lone03              0.102982   0.052896   1.947 0.051750 .
happy               0.039018   0.038299   1.019 0.308494
lifeSat             0.109931   0.060705   1.811 0.070372 .
MLQ                 0.064929   0.032332   2.008 0.044817 *
bor01               0.094382   0.029980   3.148 0.001678 **
bor02               0.003344   0.031220   0.107 0.914710
bor03               0.018609   0.025787   0.722 0.470649
consp01            -0.049076   0.021219  -2.313 0.020874 *
consp02            -0.035843   0.023006  -1.558 0.119467
consp03             0.012170   0.018825   0.646 0.518095
rankOrdLife_1       0.090538   0.035537   2.548 0.010951 *
rankOrdLife_2       0.134876   0.039049   3.454 0.000569 ***
rankOrdLife_3       0.102948   0.040412   2.547 0.010958 *
rankOrdLife_4       0.067423   0.045503   1.482 0.138640
rankOrdLife_5       0.122247   0.039451   3.099 0.001983 **
rankOrdLife_6             NA         NA      NA       NA
c19perBeh01         0.039476   0.041304   0.956 0.339372
c19perBeh02        -0.013470   0.053989  -0.249 0.803018
c19perBeh03         0.046650   0.030243   1.543 0.123178
c19RCA01            0.120630   0.024021   5.022 5.79e-07 ***
c19RCA02            0.026382   0.035172   0.750 0.453330
c19RCA03            0.032869   0.025258   1.301 0.193365
coronaClose_1      -0.179933   0.267924  -0.672 0.501962
coronaClose_2       0.121454   0.194334   0.625 0.532089
coronaClose_3       0.234169   0.184957   1.266 0.205699
coronaClose_4      -0.413784   0.157548  -2.626 0.008725 **
coronaClose_5      -0.392819   0.162865  -2.412 0.015998 *
coronaClose_6      -0.303484   0.171068  -1.774 0.076274 .
gender             -0.053925   0.082486  -0.654 0.513381
age                 0.055855   0.038644   1.445 0.148578
edu                 0.115379   0.030312   3.806 0.000147 ***
c19ProSo01          0.182293   0.031950   5.706 1.42e-08 ***
c19ProSo03          0.086675   0.029613   2.927 0.003480 **
c19ProSo04          0.107079   0.033031   3.242 0.001216 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.443 on 1383 degrees of freedom
Multiple R-squared:  0.2912,    Adjusted R-squared:  0.2661
F-statistic:  11.6 on 49 and 1383 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo03 ~ ., data = netherlands_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6555 -0.6934  0.1345  0.8374  5.4437

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.558324   0.596331   0.936  0.34930
employstatus_1       0.010323   0.130759   0.079  0.93709
employstatus_2      -0.080760   0.138489  -0.583  0.55989
employstatus_3       0.014534   0.153079   0.095  0.92437
employstatus_4      -0.148704   0.182491  -0.815  0.41529
employstatus_5      -0.118770   0.238275  -0.498  0.61824
employstatus_6      -0.126551   0.155111  -0.816  0.41471
employstatus_7      -0.044831   0.177024  -0.253  0.80011
employstatus_8      -0.278808   0.187246  -1.489  0.13672
employstatus_9      -0.322795   0.145231  -2.223  0.02640 *
employstatus_10      0.149571   0.171087   0.874  0.38214
isoFriends_inPerson  0.003456   0.018411   0.188  0.85111
isoOthPpl_inPerson   0.004021   0.017439   0.231  0.81767
isoFriends_online    0.004032   0.016080   0.251  0.80207
isoOthPpl_online     0.016899   0.015610   1.083  0.27919
lone01               0.026658   0.052152   0.511  0.60932
lone02              -0.037674   0.044270  -0.851  0.39491
lone03              -0.009857   0.047948  -0.206  0.83716
happy               -0.032973   0.034671  -0.951  0.34176
lifeSat             -0.038537   0.055007  -0.701  0.48368
MLQ                  0.003503   0.029310   0.119  0.90490
bor01                0.022339   0.027230   0.820  0.41214
bor02               -0.014219   0.028259  -0.503  0.61492
bor03                0.012156   0.023346   0.521  0.60266
consp01              0.011851   0.019242   0.616  0.53806
consp02             -0.015992   0.020840  -0.767  0.44299
consp03              0.015547   0.017039   0.912  0.36169
rankOrdLife_1        0.019870   0.032241   0.616  0.53779
rankOrdLife_2        0.038170   0.035486   1.076  0.28227
rankOrdLife_3        0.004603   0.036668   0.126  0.90012
rankOrdLife_4       -0.023327   0.041218  -0.566  0.57154
rankOrdLife_5        0.038146   0.035821   1.065  0.28710
rankOrdLife_6             NA         NA      NA       NA
c19perBeh01         -0.063693   0.037363  -1.705  0.08848 .
c19perBeh02         -0.025454   0.048869  -0.521  0.60255
c19perBeh03         -0.006497   0.027400  -0.237  0.81259
c19RCA01            -0.019533   0.021936  -0.890  0.37338
c19RCA02            -0.048655   0.031819  -1.529  0.12646
c19RCA03            -0.007515   0.022877  -0.328  0.74259
coronaClose_1        0.063399   0.242567   0.261  0.79385
coronaClose_2        0.094612   0.175924   0.538  0.59080
coronaClose_3        0.192964   0.167446   1.152  0.24936
coronaClose_4       -0.030862   0.142970  -0.216  0.82913
coronaClose_5        0.031802   0.147738   0.215  0.82959
coronaClose_6        0.095646   0.155011   0.617  0.53732
gender              -0.004302   0.074680  -0.058  0.95407
age                 -0.177650   0.034681  -5.122 3.44e-07 ***
edu                  0.006573   0.027582   0.238  0.81169
c19ProSo01           0.395714   0.027258  14.518  < 2e-16 ***
c19ProSo02           0.071026   0.024267   2.927  0.00348 **
c19ProSo04           0.465934   0.027274  17.083  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.306 on 1383 degrees of freedom
Multiple R-squared:  0.4557,    Adjusted R-squared:  0.4365
F-statistic: 23.63 on 49 and 1383 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo04 ~ ., data = netherlands_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7306 -0.5712  0.0652  0.7138  3.9733

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -1.5817030  0.5327249  -2.969 0.003038 **
employstatus_1     -0.0639955  0.1171342  -0.546 0.584918
employstatus_2      0.2895044  0.1238429   2.338 0.019546 *
employstatus_3      0.1798784  0.1370581   1.312 0.189594
employstatus_4      0.1150620  0.1635030   0.704 0.481719
employstatus_5      0.0376239  0.2134863   0.176 0.860135
employstatus_6      0.4351229  0.1385035   3.142 0.001716 **
employstatus_7      0.1136192  0.1585696   0.717 0.473788
employstatus_8      0.2475338  0.1677555   1.476 0.140289
employstatus_9      0.4658031  0.1297412   3.590 0.000342 ***
employstatus_10     0.1948442  0.1532287   1.272 0.203732
isoFriends_inPerson 0.0250816  0.0164805   1.522 0.128264
isoOthPpl_inPerson -0.0248630  0.0156096  -1.593 0.111432
isoFriends_online   0.0009586  0.0144066   0.067 0.946959
isoOthPpl_online    0.0186488  0.0139821   1.334 0.182502
lone01             -0.0436162  0.0467122  -0.934 0.350610
lone02              0.1085434  0.0395639   2.743 0.006157 **
lone03              0.0282187  0.0429502   0.657 0.511284
happy               0.0500002  0.0310429   1.611 0.107478
lifeSat             0.0736436  0.0492498   1.495 0.135062
MLQ                -0.0436963  0.0262329  -1.666 0.095998 .
bor01               0.0110654  0.0243991   0.454 0.650248
bor02              -0.0329554  0.0253035  -1.302 0.192995
bor03               0.0174828  0.0209120   0.836 0.403290
consp01            -0.0139703  0.0172372  -0.810 0.417808
consp02            -0.0185305  0.0186677  -0.993 0.321056
consp03            -0.0167981  0.0152629  -1.101 0.271271
rankOrdLife_1       0.0604308  0.0288425   2.095 0.036335 *
rankOrdLife_2       0.0545417  0.0317710   1.717 0.086255 .
rankOrdLife_3       0.0629290  0.0328071   1.918 0.055297 .
rankOrdLife_4       0.0460600  0.0369109   1.248 0.212291
rankOrdLife_5       0.0352312  0.0320913   1.098 0.272465
rankOrdLife_6             NA         NA      NA       NA
c19perBeh01         0.0946817  0.0334117   2.834 0.004667 **
c19perBeh02         0.1332789  0.0436392   3.054 0.002300 **
c19perBeh03         0.1236735  0.0243215   5.085 4.18e-07 ***
c19RCA01           -0.0131581  0.0196548  -0.669 0.503314
c19RCA02            0.0491784  0.0284997   1.726 0.084647 .
c19RCA03            0.0285018  0.0204822   1.392 0.164286
coronaClose_1       0.1491384  0.2172829   0.686 0.492589
coronaClose_2       0.0424727  0.1576220   0.269 0.787616
coronaClose_3      -0.2303493  0.1499579  -1.536 0.124744
coronaClose_4       0.1175445  0.1280496   0.918 0.358800
coronaClose_5       0.1371872  0.1323086   1.037 0.299976
coronaClose_6      -0.0529660  0.1388854  -0.381 0.702991
gender             -0.0290178  0.0669013  -0.434 0.664545
age                 0.0098482  0.0313625   0.314 0.753558
edu                -0.0063475  0.0247105  -0.257 0.797314
c19ProSo01          0.0546791  0.0261734   2.089 0.036881 *
c19ProSo02          0.0704274  0.0217252   3.242 0.001216 **
c19ProSo03          0.3739725  0.0218913  17.083  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.17 on 1383 degrees of freedom
Multiple R-squared:  0.442,     Adjusted R-squared:  0.4222
F-statistic: 22.36 on 49 and 1383 DF,  p-value: < 2.2e-16
```

(c) Repeat Question 2(b) for the other countries as a group. Which attributes are the strongest predictors? How do these attributes compare to those of your focus country?

c19ProSo01		
Best Predictors	bor03, gender, employstatus_10, isoOthPpl_inPerson, lone01, lone02, employstatus_4, employstatus_7, isoFriends_online, MLQ, c19perBeh01, c19perBeh02, c19perBeh03, age	
R-squared: 0.3535	RSE: 1.173	
c19ProSo02		
Best Predictors	employstatus_4, employstatus_5, employstatus_8, isoFriends_inPerson, isoFriends_online, lone01, lone02, lifeSat, MLQ, bor01, consp01, consp02, c19perBeh01, c19perBeh03, c19RCA01, c19RCA03, coronaClose_5, gender, edu	
R-squared: 0.394	RSE: 1.279	
c19ProSo03		
Best Predictors	employstatus_7, isoFriends_online, isoOthPpl_online, lone03, lifeSat, consp02, c19RCA03, coronaClose_1, age, edu	
R-squared: 0.454	RSE: 1.232	
c19ProSo04		
Best Predictors	employstatus_3, employstatus_8, isoFriends_inPerson, isoFriends_online, isoOthPpl_online, lone01, lone02, lifeSat, bor01, bor02, bor03, consp01, c19perBeh01, c19perBeh02, c19perBeh03, c19RCA02, c19RCA03, coronaClose_ 3, age	
R-squared: 0.3237	RSE: 1.274	

The table above summarizes the best predictors for all pro-social attitudes in other countries based on the linear regression analysis. Notably, all the best predictors identified for our focus country are also found to be the best predictors for other countries. Additionally, a greater number of predictors with extremely low p-values are observed in other countries compared to our focus country. The predictors for other countries capture a wider array of factors, suggesting a more diverse set of influences on pro-social attitudes that may vary across different cultural and societal contexts. The overall 4 models have a significantly small p-values which is <2.2e-16, indicating strong evidence against the null hypothesis and strong associations between predictors and the response variable.

The average R-squared value for other countries is similar to that of our focus country, Netherlands, suggesting a similar level of explained variability in the response variable. However, it's important to note that this level of explanation may not capture all the complexities of the underlying relationships in the data.

Moreover, the average residual standard error (RSE) for other countries is slightly smaller than that of our focus country, indicating a slightly better fit of the model to the data. This suggests that the model's predictions for pro-social attitudes in other countries are generally more accurate.

```
Call:
lm(formula = c19ProSo01 ~ ., data = other_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7147 -0.6060  0.1367  0.7187  4.5619

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -4.9640040  3.4897669  -1.422 0.154906
employstatus_1       0.0119515  0.0233089   0.513 0.608134
employstatus_2       0.0286590  0.0241884   1.185 0.236097
employstatus_3       0.0339528  0.0230706   1.472 0.141114
employstatus_4       0.1232938  0.0267776   4.604 4.15e-06 ***
employstatus_5      -0.0366610  0.0312791  -1.172 0.241181
employstatus_6      -0.0556568  0.0264881  -2.101 0.035631 *
employstatus_7      -0.1199438  0.0319555  -3.753 0.000175 ***
employstatus_8      -0.0357137  0.0526261  -0.679 0.497377
employstatus_9       0.0553469  0.0228501   2.422 0.015433 *
employstatus_10      0.2598279  0.0436397   5.954 2.64e-09 ***
isoFriends_inPerson  0.0026173  0.0029931   0.874 0.381885
isoOthPpl_inPerson   0.0181748  0.0034387   5.285 1.26e-07 ***
isoFriends_online    0.0100179  0.0030109   3.327 0.000878 ***
isoOthPpl_online     0.0047457  0.0027515   1.725 0.084583 .
lone01               0.0416548  0.0081291   5.124 3.00e-07 ***
lone02              -0.0246961  0.0071157  -3.471 0.000520 ***
lone03               0.0003060  0.0077375   0.040 0.968452
happy                0.0140480  0.0045890   3.061 0.002206 **
lifeSat             -0.0159692  0.0079960  -1.997 0.045817 *
MLQ                  0.0494756  0.0052999   9.335  < 2e-16 ***
bor01                0.0147015  0.0046107   3.189 0.001431 **
bor02               -0.0031590  0.0046065  -0.686 0.492857
bor03                0.0274327  0.0042767   6.415 1.43e-10 ***
consp01              0.0106632  0.0033658   3.168 0.001536 **
consp02             -0.0005443  0.0035401  -0.154 0.877813
consp03              0.0075452  0.0026690   2.827 0.004702 **
rankOrdLife_1        0.2407259  0.1662127   1.448 0.147542
rankOrdLife_2        0.2295132  0.1661756   1.381 0.167243
rankOrdLife_3        0.2318451  0.1662038   1.395 0.163042
rankOrdLife_4        0.2200756  0.1662762   1.324 0.185660
rankOrdLife_5        0.2257601  0.1662004   1.358 0.174358
rankOrdLife_6        0.1946721  0.1661715   1.172 0.241401
c19perBeh01          0.0738867  0.0079466   9.298  < 2e-16 ***
c19perBeh02          0.0348495  0.0095495   3.649 0.000263 ***
c19perBeh03         -0.0334529  0.0054820  -6.102 1.06e-09 ***
c19RCA01             0.0206280  0.0041243   5.002 5.72e-07 ***
c19RCA02            -0.0059957  0.0067317  -0.891 0.373110
c19RCA03            -0.0113979  0.0044629  -2.554 0.010656 *
coronaClose_1       -0.0239766  0.0659565  -0.364 0.716218
coronaClose_2        0.0301341  0.0428729   0.703 0.482141
coronaClose_3        0.0887665  0.0383090   2.317 0.020503 *
coronaClose_4        0.0915802  0.0312293   2.933 0.003365 **
coronaClose_5        0.0776760  0.0315553   2.462 0.013838 *
coronaClose_6        0.0230664  0.0324148   0.712 0.476716
gender               0.0878008  0.0136831   6.417 1.41e-10 ***
age                  0.0238315  0.0058059   4.105 4.06e-05 ***
edu                 -0.0004197  0.0048656  -0.086 0.931261
c19ProSo02           0.2096662  0.0048755  43.004  < 2e-16 ***
c19ProSo03           0.2641896  0.0049955  52.885  < 2e-16 ***
c19ProSo04           0.1105916  0.0049929  22.150  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.173 on 33531 degrees of freedom
Multiple R-squared:  0.3545,    Adjusted R-squared:  0.3535
F-statistic: 368.2 on 50 and 33531 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo02 ~ ., data = other_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9125 -0.6779  0.1441  0.7940  5.3710

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.982227   3.805449   0.258 0.796323
employstatus_1      -0.027847   0.025416  -1.096 0.273248
employstatus_2      -0.086728   0.026372  -3.289 0.001008 **
employstatus_3       0.003338   0.025158   0.133 0.894441
employstatus_4      -0.238117   0.029179  -8.160 3.45e-16 ***
employstatus_5      -0.120049   0.034102  -3.520 0.000432 ***
employstatus_6       0.035830   0.028885   1.240 0.214812
employstatus_7      -0.019895   0.034852  -0.571 0.568109
employstatus_8      -0.249461   0.057369  -4.348 1.38e-05 ***
employstatus_9       0.015013   0.024918   0.602 0.546862
employstatus_10      0.034128   0.047611   0.717 0.473491
isoFriends_inPerson  0.021833   0.003262   6.694 2.21e-11 ***
isoOthPpl_inPerson  -0.012069   0.003751  -3.218 0.001292 **
isoFriends_online    0.015209   0.003283   4.633 3.62e-06 ***
isoOthPpl_online     0.009310   0.003000   3.103 0.001916 **
lone01               0.030862   0.008866   3.481 0.000500 ***
lone02              -0.027694   0.007759  -3.569 0.000358 ***
lone03               0.005366   0.008437   0.636 0.524761
happy                0.015827   0.005004   3.163 0.001563 **
lifeSat              0.054072   0.008715   6.205 5.54e-10 ***
MLQ                  0.064432   0.005776  11.155  < 2e-16 ***
bor01                0.042780   0.005023   8.517  < 2e-16 ***
bor02               -0.007657   0.005023  -1.524 0.127426
bor03               -0.002556   0.004666  -0.548 0.583807
consp01             -0.016840   0.003670  -4.589 4.47e-06 ***
consp02             -0.014287   0.003859  -3.702 0.000214 ***
consp03              0.004168   0.002911   1.432 0.152201
rankOrdLife_1       -0.074425   0.181248  -0.411 0.681350
rankOrdLife_2       -0.041412   0.181207  -0.229 0.819231
rankOrdLife_3       -0.080519   0.181238  -0.444 0.656849
rankOrdLife_4       -0.058885   0.181317  -0.325 0.745364
rankOrdLife_5       -0.055968   0.181234  -0.309 0.757462
rankOrdLife_6       -0.086609   0.181201  -0.478 0.632673
c19perBeh01          0.035503   0.008674   4.093 4.27e-05 ***
c19perBeh02          0.015782   0.010415   1.515 0.129689
c19perBeh03          0.023783   0.005980   3.977 6.98e-05 ***
c19RCA01             0.067819   0.004484  15.126  < 2e-16 ***
c19RCA02            -0.010142   0.007340  -1.382 0.167095
c19RCA03             0.067599   0.004853  13.929  < 2e-16 ***
coronaClose_1       -0.166012   0.071915  -2.308 0.020980 *
coronaClose_2        0.082120   0.046748   1.757 0.078984 .
coronaClose_3       -0.039194   0.041776  -0.938 0.348155
coronaClose_4        0.022990   0.034057   0.675 0.499666
coronaClose_5        0.145158   0.034403   4.219 2.46e-05 ***
coronaClose_6       -0.003887   0.035346  -0.110 0.912440
gender              -0.062515   0.014926  -4.188 2.82e-05 ***
age                 -0.020135   0.006332  -3.180 0.001473 **
edu                  0.043140   0.005300   8.139 4.12e-16 ***
c19ProSo01           0.249300   0.005797  43.004  < 2e-16 ***
c19ProSo03           0.345124   0.005347  64.540  < 2e-16 ***
c19ProSo04           0.031275   0.005481   5.706 1.17e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.279 on 33531 degrees of freedom
Multiple R-squared:  0.394,     Adjusted R-squared:  0.3931
F-statistic: 436.1 on 50 and 33531 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo03 ~ ., data = other_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3540 -0.6905  0.1703  0.7247  6.1031

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.884e+00  3.665e+00   0.514 0.607272
employstatus_1       3.550e-02  2.448e-02   1.450 0.147065
employstatus_2       1.902e-02  2.540e-02   0.749 0.453961
employstatus_3       4.170e-02  2.423e-02   1.721 0.085230 .
employstatus_4       3.425e-02  2.813e-02   1.217 0.223477
employstatus_5      -1.986e-02  3.285e-02  -0.604 0.545585
employstatus_6      -2.746e-02  2.782e-02  -0.987 0.323704
employstatus_7      -1.189e-01  3.356e-02  -3.543 0.000397 ***
employstatus_8      -1.297e-02  5.527e-02  -0.235 0.814413
employstatus_9       3.288e-02  2.400e-02   1.370 0.170644
employstatus_10      9.319e-02  4.585e-02   2.032 0.042143 *
isoFriends_inPerson  9.050e-03  3.143e-03   2.879 0.003989 **
isoOthPpl_inPerson   8.582e-03  3.613e-03   2.375 0.017538 *
isoFriends_online   -1.055e-02  3.162e-03  -3.335 0.000854 ***
isoOthPpl_online     1.468e-02  2.889e-03   5.081 3.77e-07 ***
lone01              -6.540e-04  8.541e-03  -0.077 0.938960
lone02              -7.154e-03  7.475e-03  -0.957 0.338493
lone03               3.946e-02  8.124e-03   4.858 1.19e-06 ***
happy                2.681e-03  4.820e-03   0.556 0.578147
lifeSat              3.526e-02  8.396e-03   4.199 2.68e-05 ***
MLQ                 -1.364e-02  5.573e-03  -2.447 0.014394 *
bor01               -8.713e-04  4.843e-03  -0.180 0.857238
bor02                1.212e-02  4.838e-03   2.505 0.012259 *
bor03                1.249e-02  4.494e-03   2.780 0.005440 **
consp01             -8.010e-03  3.535e-03  -2.266 0.023469 *
consp02             -1.713e-02  3.717e-03  -4.608 4.08e-06 ***
consp03              4.682e-03  2.803e-03   1.670 0.094915 .
rankOrdLife_1       -1.096e-01  1.746e-01  -0.628 0.530275
rankOrdLife_2       -7.261e-02  1.745e-01  -0.416 0.677383
rankOrdLife_3       -1.082e-01  1.746e-01  -0.620 0.535388
rankOrdLife_4       -1.105e-01  1.746e-01  -0.633 0.526973
rankOrdLife_5       -8.889e-02  1.746e-01  -0.509 0.610609
rankOrdLife_6       -1.252e-01  1.745e-01  -0.717 0.473194
c19perBeh01         -4.232e-05  8.357e-03  -0.005 0.995959
c19perBeh02         -2.668e-02  1.003e-02  -2.660 0.007815 **
c19perBeh03          1.012e-02  5.760e-03   1.758 0.078829 .
c19RCA01             6.752e-03  4.333e-03   1.558 0.119193
c19RCA02            -1.905e-02  7.069e-03  -2.695 0.007034 **
c19RCA03            -2.517e-02  4.686e-03  -5.371 7.89e-08 ***
coronaClose_1        2.429e-01  6.926e-02   3.507 0.000453 ***
coronaClose_2       -2.616e-03  4.503e-02  -0.058 0.953671
coronaClose_3        3.018e-02  4.024e-02   0.750 0.453221
coronaClose_4       -9.007e-03  3.280e-02  -0.275 0.783634
coronaClose_5       -7.516e-02  3.314e-02  -2.268 0.023341 *
coronaClose_6       -4.882e-02  3.404e-02  -1.434 0.151589
gender               2.108e-02  1.438e-02   1.466 0.142717
age                 -5.255e-02  6.093e-03  -8.625  < 2e-16 ***
edu                  2.055e-02  5.109e-03   4.022 5.79e-05 ***
c19ProSo01           2.914e-01  5.510e-03  52.885  < 2e-16 ***
c19ProSo02           3.202e-01  4.961e-03  64.540  < 2e-16 ***
c19ProSo04           3.015e-01  5.019e-03  60.072  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.232 on 33531 degrees of freedom
Multiple R-squared:  0.454,     Adjusted R-squared:  0.4532
F-statistic: 557.7 on 50 and 33531 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo04 ~ ., data = other_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9914 -0.6696  0.1317  0.8007  5.5698

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -3.9211289  3.7894392  -1.035 0.300792
employstatus_1       0.0070658  0.0253102   0.279 0.780117
employstatus_2       0.0747525  0.0262626   2.846 0.004425 **
employstatus_3       0.1081735  0.0250452   4.319 1.57e-05 ***
employstatus_4       0.0805033  0.0290825   2.768 0.005641 **
employstatus_5       0.0445281  0.0339644   1.311 0.189859
employstatus_6      -0.0070681  0.0287642  -0.246 0.805896
employstatus_7       0.0450739  0.0347055   1.299 0.194038
employstatus_8       0.2381063  0.0571300   4.168 3.08e-05 ***
employstatus_9      -0.0354051  0.0248133  -1.427 0.153630
employstatus_10      0.1388039  0.0474054   2.928 0.003414 **
isoFriends_inPerson -0.0141416  0.0032492  -4.352 1.35e-05 ***
isoOthPpl_inPerson   0.0007431  0.0037355   0.199 0.842310
isoFriends_online    0.0143371  0.0032690   4.386 1.16e-05 ***
isoOthPpl_online    -0.0135392  0.0029870  -4.533 5.84e-06 ***
lone01              -0.0313889  0.0088288  -3.555 0.000378 ***
lone02               0.0429832  0.0077245   5.565 2.65e-08 ***
lone03               0.0290380  0.0084004   3.457 0.000547 ***
happy               -0.0107097  0.0049834  -2.149 0.031635 *
lifeSat              0.0349940  0.0086809   4.031 5.56e-05 ***
MLQ                 -0.0130401  0.0057620  -2.263 0.023633 *
bor01               -0.0187037  0.0050063  -3.736 0.000187 ***
bor02                0.0244877  0.0050002   4.897 9.76e-07 ***
bor03                0.0237933  0.0046449   5.122 3.03e-07 ***
consp01              0.0263229  0.0036525   7.207 5.85e-13 ***
consp02             -0.0075419  0.0038438  -1.962 0.049759 *
consp03             -0.0080104  0.0028982  -2.764 0.005714 **
rankOrdLife_1        0.2141665  0.1804850   1.187 0.235387
rankOrdLife_2        0.1711589  0.1804455   0.949 0.342864
rankOrdLife_3        0.2062762  0.1804751   1.143 0.253063
rankOrdLife_4        0.1504522  0.1805549   0.833 0.404694
rankOrdLife_5        0.1449951  0.1804730   0.803 0.421739
rankOrdLife_6        0.1345446  0.1804406   0.746 0.455887
c19perBeh01          0.0306329  0.0086383   3.546 0.000391 ***
c19perBeh02          0.1484237  0.0103397  14.355  < 2e-16 ***
c19perBeh03          0.0756333  0.0059416  12.729  < 2e-16 ***
c19RCA01             0.0130156  0.0044795   2.906 0.003668 **
c19RCA02             0.1086267  0.0072856  14.910  < 2e-16 ***
c19RCA03            -0.0426617  0.0048410  -8.813  < 2e-16 ***
coronaClose_1        0.1337413  0.0716157   1.867 0.061843 .
coronaClose_2        0.1097995  0.0465502   2.359 0.018343 *
coronaClose_3        0.1906645  0.0415884   4.585 4.57e-06 ***
coronaClose_4        0.0725613  0.0339126   2.140 0.032390 *
coronaClose_5       -0.0666728  0.0342657  -1.946 0.051692 .
coronaClose_6       -0.0465095  0.0351971  -1.321 0.186377
gender              -0.0174816  0.0148667  -1.176 0.239647
age                  0.0477888  0.0063006   7.585 3.42e-14 ***
edu                  0.0056104  0.0052832   1.062 0.288275
c19ProSo01           0.1303967  0.0058870  22.150  < 2e-16 ***
c19ProSo02           0.0310129  0.0054355   5.706 1.17e-08 ***
c19ProSo03           0.3222730  0.0053648  60.072  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.274 on 33531 degrees of freedom
Multiple R-squared:  0.3237,    Adjusted R-squared:  0.3227
F-statistic:   321 on 50 and 33531 DF,  p-value: < 2.2e-16
```

3. Focus country vs cluster of similar countries. (10 Marks)

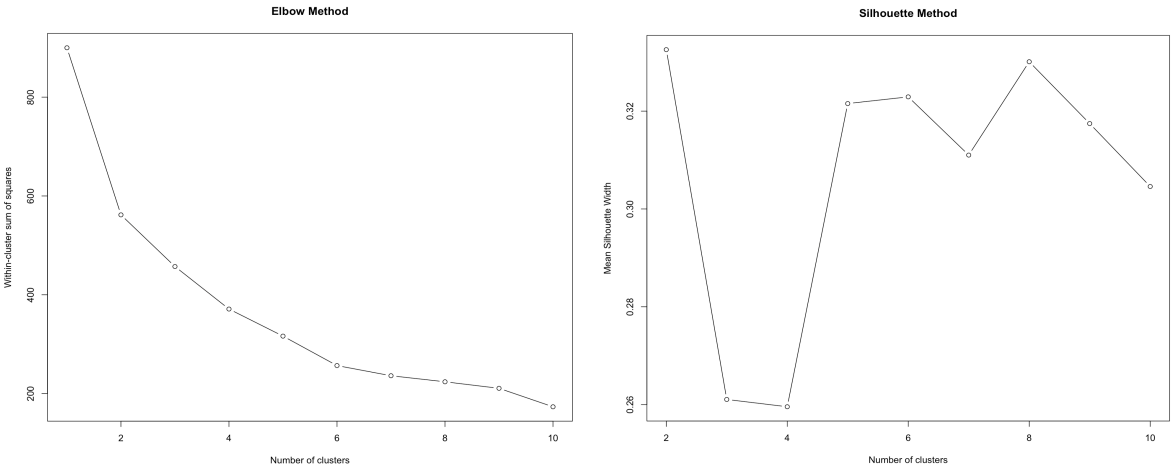
(a) Using a collection of social, economic, health, political or other indicators from external sources, identify at least 5 countries in the baseline data that are similar to your focus country using clustering. Van Lissa (2022) refers to several indicators you might consider, among others. Some of these are listed in the references, but these are not exhaustive. State the indicators used and describe how you calculated/identified similar countries. Copy and paste the table of values you used for your clustering into your report as an Appendix.

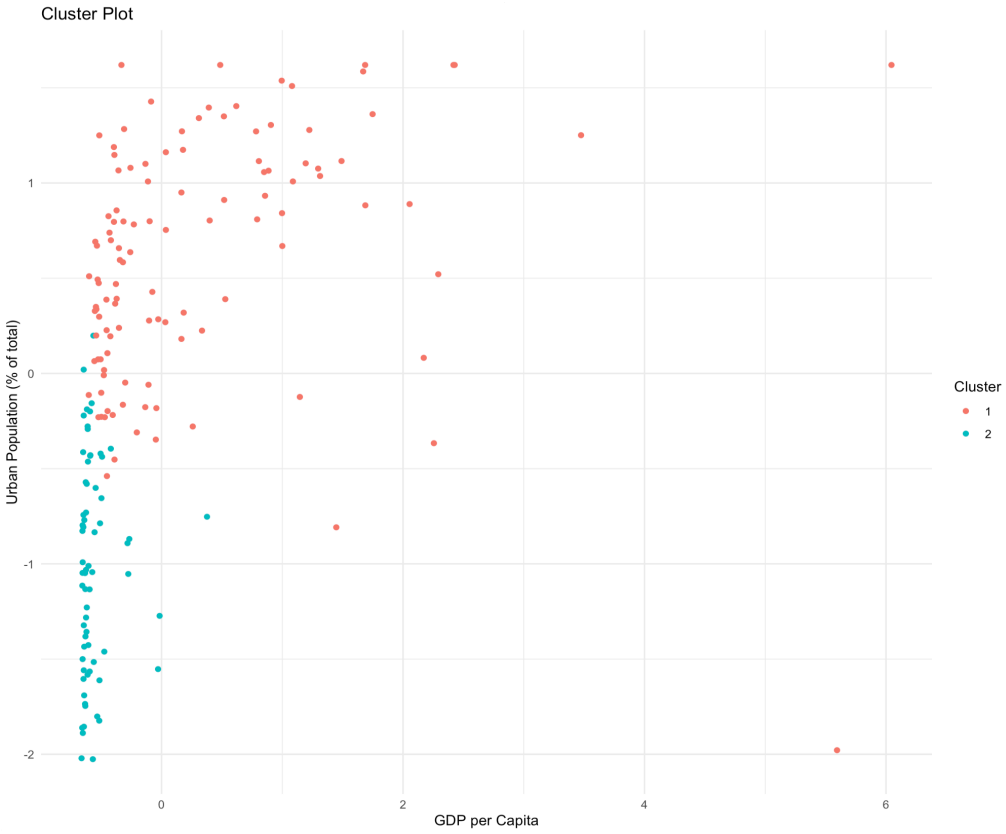
Indicator	Variable Name	Description
GDP per capita (current US$)	latest.value_NY.GDP.PCAP.CD	This indicator measures the average economic output per person in a country, calculated by dividing the country's gross domestic product (GDP) by its population.
Urban Population (% of total population)	latest.value_SP.URB.TOTL.IN.ZS	This indicator represents the percentage of a country's total population that resides in urban areas.
Rural population (% of total population)	latest.value_SP.RUR.TOTL.ZS	This indicator represents the percentage of a country's total population that resides in rural areas.
Birth rate	latest.value_birthrate.2054	The birth rate refers to the number of live births per 1,000 people in a given population within a specified period, usually one year.
Death rate	latest.value_deathrate.2066	The death rate, also known as the mortality rate, refers to the number of deaths per 1,000 people in a given population within a specified period, usually one year.

These indicators have been chosen for their relevance in assessing the economic and demographic characteristics of countries, aiding in the identification of similar nations to our focus country. GDP per capita serves as a key measure of economic prosperity and living standards, while the percentage of urban and rural populations sheds light on a country's urbanization and agricultural landscape, respectively. Birth and death rates provide insights into demographic trends, fertility patterns, and overall population health.

To determine the optimal number of clusters for our analysis, I utilized the elbow and silhouette methods. These techniques pointed towards 2 as the ideal number of clusters. Subsequently, I conducted k-means clustering with this optimal value and identified countries sharing similarities with our focus country, Netherlands. I calculated distances between each country which belongs to the same cluster as Netherlands using Euclidean distance and sorted the data frame based on distances and get the closest countries with Netherlands.

The 6 countries identified as similar to Netherlands include **Sweden**, **Greenland**, **Austria**, **Australia**, **Finland** and **San Marino**, all falling within the same cluster. These countries exhibit comparable economic and demographic characteristics, forming a cluster that shares resemblances with Netherlands in terms of the selected indicators.





(b) How well do participant responses (attributes) predict pro-social attitudes **(c19ProSo01,2,3 and 4)** for this cluster of similar countries? Which attributes are the strongest predictors? How do these attributes compare to those of your focus country? Comment on the similarity and/or difference between your results for this question and Question 2(c). That is, does the group of all other countries 2(c), or the cluster of similar countries 3(b) give a better match to the important attributes for predicting pro-social attitudes in your focus country? Discuss.

Similar countries (Sweden, Greenland, Austria, Australia, Finland, San Marino)	**c19ProSo01**		
	Best Predictors	c19perBeh01, p-value: 0.00116	
		bor03, p-value: 0. 00117	
	R-squared: 0.4398	RSE: 1.134	
	c19ProSo02		
	Best Predictors	c19perBeh01, p-value: 0.00052	
		MLQ, p-value: 0.00303	
	R-squared: 0.4809	RSE: 1.183	
	c19ProSo03		
	Best Predictors	age, p-value: 0.00563	
	R-squared: 0.5432	RSE: 1.156	
	c19ProSo04		
	Best Predictors	consp03, p-value: 0.000146	
		c19perBeh02, p-value: 0.000255	
	R-squared: 0.4152	RSE: 1.075	

The table above are the best predictors for pro-social attitudes in similar countries. These 4 models for similar countries have an extremely small p-value which is <2.2e-16, indicating that there is strong evidence against the null hypothesis where the independent variables in the model are strongly associated with the pro-social attitude attributes. The average R-squared value for similar countries is higher than our focus country which indicates that this model has more usefulness than Netherlands, but it may not capture all the underlying relationships in the data. The average RSE for similar countries is lower than Netherlands, indicating these models fit the data better. This suggests that the predictions generated by these models are generally closer to the actual values, enhancing their reliability.

The predictors for similar countries are slightly different than the predictors for Netherlands. Differences in predictors could stem from other various factors such as differences in education systems, government policies and many more. Therefore, the unique characteristics and dynamics of each country could contribute to differences in predictors observed between the Netherlands and other similar countries. Interestingly, we can also see that for Netherlands and its similar countries have common best predictor such as corona personal behavior (c19perBeh01,02,03), indicating that these countries have common understanding of the importance of individual and collective actions in mitigating the impact of the coronavirus pandemic. Similar countries often share similar cultural, socioeconomic, and contextual backgrounds. These similarities can lead to common challenges, priorities, and behaviors among their populations.

When comparing between similar countries and other countries, it seems that the predictors for similar countries gave a more focused and specific understanding of pro-social attitudes. This may be attributed to their similar cultural and socioeconomic backgrounds. While the predictors for other countries capture a wider array of factors, suggesting a more diverse set of influences on pro-social attitudes that may vary across different cultural and societal contexts. It's worth noting that while many predictors for similar countries overlap with those of other countries, the reverse is not true, indicating distinct patterns in the predictors for similar countries.

```
Call:
lm(formula = c19ProSo01 ~ ., data = similar_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9583 -0.5858  0.0006  0.6597  3.5840

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.0391693  0.7140324   0.055  0.95627
employstatus_1     -0.0245467  0.1785098  -0.138  0.89067
employstatus_2     -0.0228449  0.1876587  -0.122  0.90314
employstatus_3     -0.1190894  0.1977115  -0.602  0.54713
employstatus_4      0.2856486  0.2141277   1.334  0.18261
employstatus_5      0.0334801  0.2403874   0.139  0.88927
employstatus_6     -0.0787578  0.1954857  -0.403  0.68715
employstatus_7     -0.0621721  0.2192769  -0.284  0.77685
employstatus_8      0.0141530  0.2646614   0.053  0.95737
employstatus_9     -0.0166373  0.2013563  -0.083  0.93417
employstatus_10    -0.1411062  0.3782648  -0.373  0.70923
isoFriends_inPerson 0.0021007  0.0204476   0.103  0.91820
isoOthPpl_inPerson  0.0169974  0.0227685   0.747  0.45558
isoFriends_online   0.0210984  0.0193850   1.088  0.27678
isoOthPpl_online    0.0009694  0.0205672   0.047  0.96242
lone01              0.0539952  0.0612765   0.881  0.37851
lone02             -0.0775124  0.0564502  -1.373  0.17013
lone03              0.0485204  0.0587521   0.826  0.40916
happy              -0.0759035  0.0341300  -2.224  0.02645 *
lifeSat             0.0375941  0.0629342   0.597  0.55045
MLQ                 0.0307961  0.0398430   0.773  0.43981
bor01               0.0198225  0.0324544   0.611  0.54153
bor02               0.0194672  0.0335000   0.581  0.56134
bor03               0.1043753  0.0320305   3.259  0.00117 **
consp01            -0.0323420  0.0227951  -1.419  0.15637
consp02             0.0238229  0.0247439   0.963  0.33597
consp03            -0.0130798  0.0204344  -0.640  0.52232
rankOrdLife_1       0.0892775  0.0372767   2.395  0.01687 *
rankOrdLife_2       0.0598799  0.0430581   1.391  0.16474
rankOrdLife_3       0.0081216  0.0442923   0.183  0.85456
rankOrdLife_4       0.0656043  0.0440401   1.490  0.13674
rankOrdLife_5      -0.0084259  0.0425716  -0.198  0.84316
rankOrdLife_6              NA         NA      NA       NA
c19perBeh01         0.1644432  0.0504362   3.260  0.00116 **
c19perBeh02        -0.1545437  0.0611419  -2.528  0.01169 *
c19perBeh03         0.0287734  0.0351153   0.819  0.41282
c19RCA01            0.0699986  0.0298323   2.346  0.01922 *
c19RCA02            0.0104615  0.0444389   0.235  0.81395
c19RCA03           -0.0856204  0.0329662  -2.597  0.00958 **
coronaClose_1      -0.2912803  0.4220190  -0.690  0.49028
coronaClose_2      -0.3773625  0.3272622  -1.153  0.24925
coronaClose_3       0.1035821  0.3415278   0.303  0.76175
coronaClose_4      -0.3141547  0.2618083  -1.200  0.23054
coronaClose_5      -0.2072526  0.2645824  -0.783  0.43369
coronaClose_6      -0.3872678  0.2631603  -1.472  0.14155
gender              0.0331006  0.0842461   0.393  0.69450
age                 0.0338501  0.0364018   0.930  0.35273
edu                 0.0046545  0.0314164   0.148  0.88226
c19ProSo02          0.2037948  0.0343814   5.927 4.71e-09 ***
c19ProSo03          0.3283178  0.0339533   9.670  < 2e-16 ***
c19ProSo04          0.1114725  0.0385147   2.894  0.00391 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.134 on 742 degrees of freedom
Multiple R-squared:  0.4398,    Adjusted R-squared:  0.4028
F-statistic: 11.89 on 49 and 742 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo02 ~ ., data = similar_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9125 -0.5772  0.1140  0.6737  3.1569

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -1.6970630  0.7423736  -2.286  0.02254 *
employstatus_1      0.2415387  0.1860384   1.298  0.19458
employstatus_2      0.1225991  0.1957429   0.626  0.53129
employstatus_3      0.2759607  0.2060827   1.339  0.18096
employstatus_4     -0.0726773  0.2236608  -0.325  0.74531
employstatus_5     -0.0965199  0.2507851  -0.385  0.70044
employstatus_6     -0.0852892  0.2039572  -0.418  0.67594
employstatus_7     -0.1245652  0.2287481  -0.545  0.58623
employstatus_8     -0.5675887  0.2753462  -2.061  0.03962 *
employstatus_9      0.0489584  0.2100772   0.233  0.81579
employstatus_10    -0.1647970  0.3946512  -0.418  0.67638
isoFriends_inPerson -0.0005955 0.0213340  -0.028  0.97774
isoOthPpl_inPerson  0.0033055  0.0237640   0.139  0.88941
isoFriends_online   0.0078614  0.0202393   0.388  0.69781
isoOthPpl_online   -0.0072929  0.0214571  -0.340  0.73404
lone01             -0.0523761  0.0639370  -0.819  0.41295
lone02              0.0290792  0.0589621   0.493  0.62203
lone03              0.0690187  0.0612745   1.126  0.26037
happy               0.0722639  0.0356292   2.028  0.04289 *
lifeSat             0.0049902  0.0656775   0.076  0.93945
MLQ                 0.1229512  0.0413410   2.974  0.00303 **
bor01               0.0382090  0.0338405   1.129  0.25923
bor02               0.0347968  0.0349367   0.996  0.31958
bor03              -0.0183626  0.0336504  -0.546  0.58544
consp01            -0.0289713  0.0237916  -1.218  0.22372
consp02            -0.0158163  0.0258260  -0.612  0.54045
consp03             0.0462703  0.0212583   2.177  0.02983 *
rankOrdLife_1       0.0736237  0.0389488   1.890  0.05911 .
rankOrdLife_2       0.0651043  0.0449194   1.449  0.14766
rankOrdLife_3      -0.0649832  0.0461516  -1.408  0.15954
rankOrdLife_4      -0.0295222  0.0460049  -0.642  0.52125

rankOrdLife_5       0.0139762  0.0444150   0.315  0.75310
rankOrdLife_6              NA         NA      NA       NA
c19perBeh01        -0.0683903  0.0529384  -1.292  0.19680
c19perBeh02         0.0345934  0.0640535   0.540  0.58931
c19perBeh03         0.1267191  0.0363575   3.485  0.00052 ***
c19RCA01            0.0659674  0.0311467   2.118  0.03451 *
c19RCA02           -0.0333114  0.0463507  -0.719  0.47256
c19RCA03            0.0072744  0.0345501   0.211  0.83330
coronaClose_1       0.7608234  0.4395660   1.731  0.08389 .
coronaClose_2       0.4486797  0.3413559   1.314  0.18912
coronaClose_3       0.3767053  0.3560847   1.058  0.29044
coronaClose_4      -0.0619556  0.2734117  -0.227  0.82080
coronaClose_5       0.1323407  0.2761220   0.479  0.63188
coronaClose_6       0.3426280  0.2746794   1.247  0.21265
gender             -0.1058553  0.0878209  -1.205  0.22845
age                 0.0298624  0.0379860   0.786  0.43203
edu                 0.0586062  0.0327079   1.792  0.07357 .
c19ProSo01          0.2218444  0.0374265   5.927 4.71e-09 ***
c19ProSo03          0.3587217  0.0352086  10.188  < 2e-16 ***
c19ProSo04          0.0090367  0.0404090   0.224  0.82311
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.183 on 742 degrees of freedom
Multiple R-squared:  0.4809,    Adjusted R-squared:  0.4466
F-statistic: 14.03 on 49 and 742 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo03 ~ ., data = similar_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6772 -0.6185  0.0773  0.6380  4.1351

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.777217   0.726990  -1.069  0.28538
employstatus_1     -0.143966   0.181814  -0.792  0.42871
employstatus_2     -0.006939   0.191213  -0.036  0.97106
employstatus_3      0.107582   0.201464   0.534  0.59350
employstatus_4     -0.085973   0.218420  -0.394  0.69398
employstatus_5     -0.131124   0.244894  -0.535  0.59251
employstatus_6      0.127826   0.199152   0.642  0.52117
employstatus_7     -0.022637   0.223438  -0.101  0.91933
employstatus_8     -0.096719   0.269648  -0.359  0.71993
employstatus_9      0.284635   0.204902   1.389  0.16521
employstatus_10     0.275993   0.385328   0.716  0.47406
isoFriends_inPerson 0.019698   0.020822   0.946  0.34446
isoOthPpl_inPerson -0.031565   0.023179  -1.362  0.17368
isoFriends_online   0.002890   0.019767   0.146  0.88379
isoOthPpl_online    0.035544   0.020916   1.699  0.08967 .
lone01              0.011831   0.062468   0.189  0.84984
lone02              0.004749   0.057591   0.082  0.93430
lone03             -0.007435   0.059891  -0.124  0.90124
happy              -0.006864   0.034891  -0.197  0.84410
lifeSat             0.069524   0.064090   1.085  0.27837
MLQ                -0.039480   0.040588  -0.973  0.33101
bor01              -0.008633   0.033075  -0.261  0.79416
bor02              -0.066088   0.034056  -1.941  0.05269 .
bor03              -0.039982   0.032837  -1.218  0.22376
consp01             0.008675   0.023256   0.373  0.70924
consp02            -0.040252   0.025185  -1.598  0.11041
consp03             0.007301   0.020825   0.351  0.72599
rankOrdLife_1      -0.044895   0.038093  -1.179  0.23895
rankOrdLife_2       0.042014   0.043903   0.957  0.33889
rankOrdLife_3       0.086356   0.045020   1.918  0.05547 .
rankOrdLife_4       0.042343   0.044914   0.943  0.34611
rankOrdLife_5       0.056342   0.043329   1.300  0.19389
rankOrdLife_6            NA         NA      NA       NA
c19perBeh01         0.009767   0.051756   0.189  0.85038
c19perBeh02         0.023384   0.062561   0.374  0.70867
c19perBeh03        -0.060101   0.035728  -1.682  0.09296 .
c19RCA01           -0.006631   0.030509  -0.217  0.82801
c19RCA02           -0.048070   0.045247  -1.062  0.28841
c19RCA03            0.029091   0.033726   0.863  0.38866
coronaClose_1       0.079212   0.430135   0.184  0.85394
coronaClose_2       0.063441   0.333747   0.190  0.84929
coronaClose_3       0.126069   0.347983   0.362  0.71724
coronaClose_4       0.302702   0.266791   1.135  0.25691
coronaClose_5       0.199256   0.269603   0.739  0.46010
coronaClose_6      -0.066365   0.268522  -0.247  0.80486
gender              0.060651   0.085821   0.707  0.47996
age                -0.102523   0.036921  -2.777  0.00563 **
edu                 0.036662   0.031983   1.146  0.25204
c19ProSo01          0.340865   0.035251   9.670  < 2e-16 ***
c19ProSo02          0.342129   0.033580  10.188  < 2e-16 ***
c19ProSo04          0.295339   0.037946   7.783 2.38e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.156 on 742 degrees of freedom
Multiple R-squared:  0.5432,    Adjusted R-squared:  0.5131
F-statistic: 18.01 on 49 and 742 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = c19ProSo04 ~ ., data = similar_countries_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9277 -0.5831  0.0320  0.6716  3.0912

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.859842   0.676051  -1.272 0.203821
employstatus_1     -0.006507   0.169200  -0.038 0.969334
employstatus_2      0.112915   0.177823   0.635 0.525633
employstatus_3     -0.012849   0.187443  -0.069 0.945366
employstatus_4     -0.074712   0.203183  -0.368 0.713196
employstatus_5      0.152508   0.227782   0.670 0.503362
employstatus_6     -0.016836   0.185308  -0.091 0.927633
employstatus_7     -0.029754   0.207847  -0.143 0.886206
employstatus_8      0.190365   0.250759   0.759 0.448001
employstatus_9     -0.260697   0.190614  -1.368 0.171829
employstatus_10     0.508183   0.358081   1.419 0.156266
isoFriends_inPerson -0.029141   0.019352  -1.506 0.132529
isoOthPpl_inPerson  0.011695   0.021585   0.542 0.588111
isoFriends_online  -0.034104   0.018346  -1.859 0.063427 .
isoOthPpl_online    0.025165   0.019472   1.292 0.196636
lone01             -0.042158   0.058090  -0.726 0.468233
lone02              0.129690   0.053361   2.430 0.015318 *
lone03             -0.077013   0.055641  -1.384 0.166745
happy               0.020503   0.032449   0.632 0.527664
lifeSat             0.010539   0.059664   0.177 0.859844
MLQ                -0.045104   0.037744  -1.195 0.232464
bor01              -0.012019   0.030766  -0.391 0.696171
bor02               0.036852   0.031731   1.161 0.245855
bor03               0.041185   0.030539   1.349 0.177877
consp01             0.074135   0.021463   3.454 0.000584 ***
consp02            -0.027743   0.023446  -1.183 0.237082
consp03            -0.073246   0.019186  -3.818 0.000146 ***
rankOrdLife_1       0.075357   0.035360   2.131 0.033409 *
rankOrdLife_2       0.020691   0.040858   0.506 0.612710
rankOrdLife_3       0.044973   0.041950   1.072 0.284049
rankOrdLife_4       0.016072   0.041801   0.384 0.700724
rankOrdLife_5      -0.001187   0.040352  -0.029 0.976544
rankOrdLife_6            NA         NA      NA       NA
c19perBeh01         0.075254   0.048067   1.566 0.117867
c19perBeh02         0.211947   0.057679   3.675 0.000255 ***
c19perBeh03         0.047212   0.033253   1.420 0.156097
c19RCA01           -0.010066   0.028378  -0.355 0.722913
c19RCA02            0.150642   0.041758   3.608 0.000330 ***
c19RCA03           -0.018791   0.031381  -0.599 0.549491
coronaClose_1       0.418761   0.399838   1.047 0.295290
coronaClose_2       0.515352   0.309892   1.663 0.096734 .
coronaClose_3       0.241158   0.323611   0.745 0.456381
coronaClose_4       0.336697   0.248084   1.357 0.175135
coronaClose_5       0.194677   0.250783   0.776 0.437833
coronaClose_6       0.539852   0.249009   2.168 0.030476 *
gender             -0.033980   0.079850  -0.426 0.670561
age                 0.032637   0.034502   0.946 0.344493
edu                -0.030831   0.029756  -1.036 0.300494
c19ProSo01          0.100146   0.034601   2.894 0.003912 **
c19ProSo02          0.007458   0.033349   0.224 0.823105
c19ProSo03          0.255564   0.032836   7.783 2.38e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.075 on 742 degrees of freedom
Multiple R-squared:  0.4152,    Adjusted R-squared:  0.3766
F-statistic: 10.75 on 49 and 742 DF,  p-value: < 2.2e-16
```

Appendix

I acknowledge the use of ChatGPT (https://chat.openai.com/) to generate R codes and refine the language for my own work.

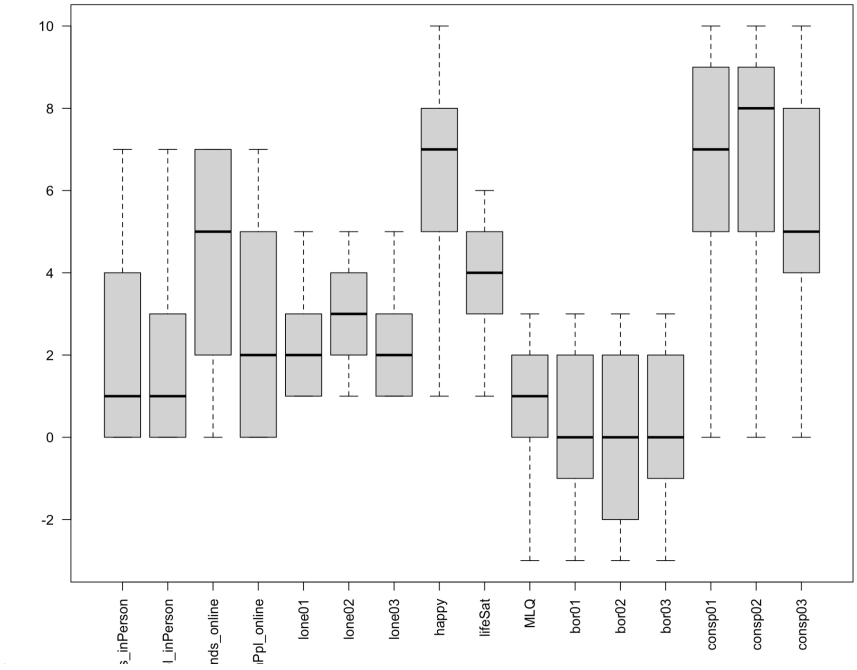


Figure 1.1.1

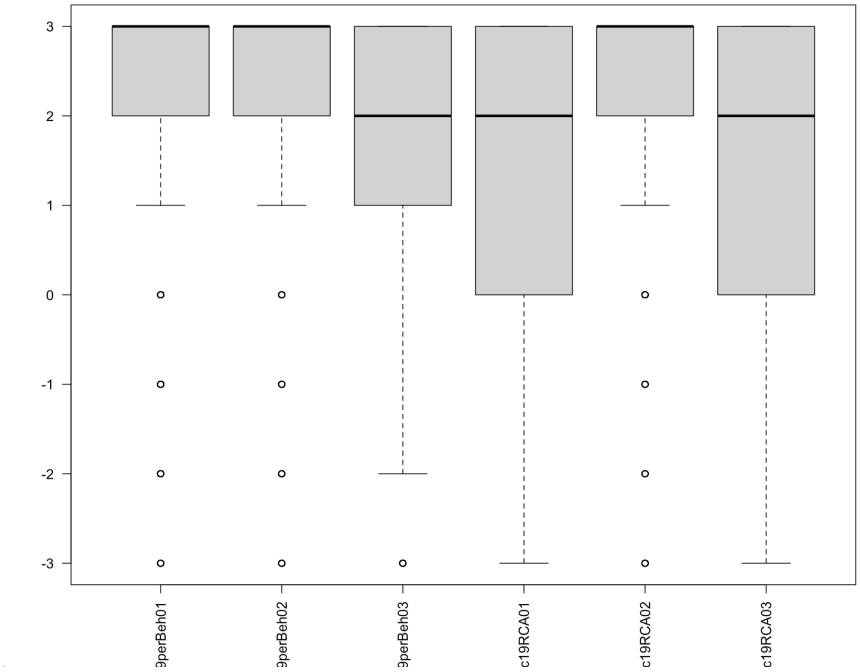


Figure 1.1.2

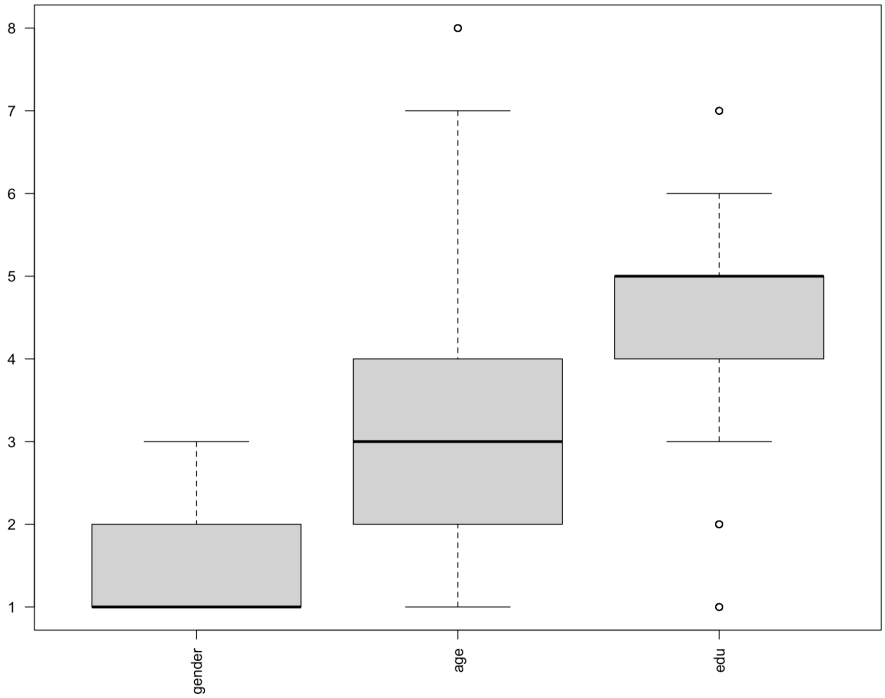


Figure 1.1.3

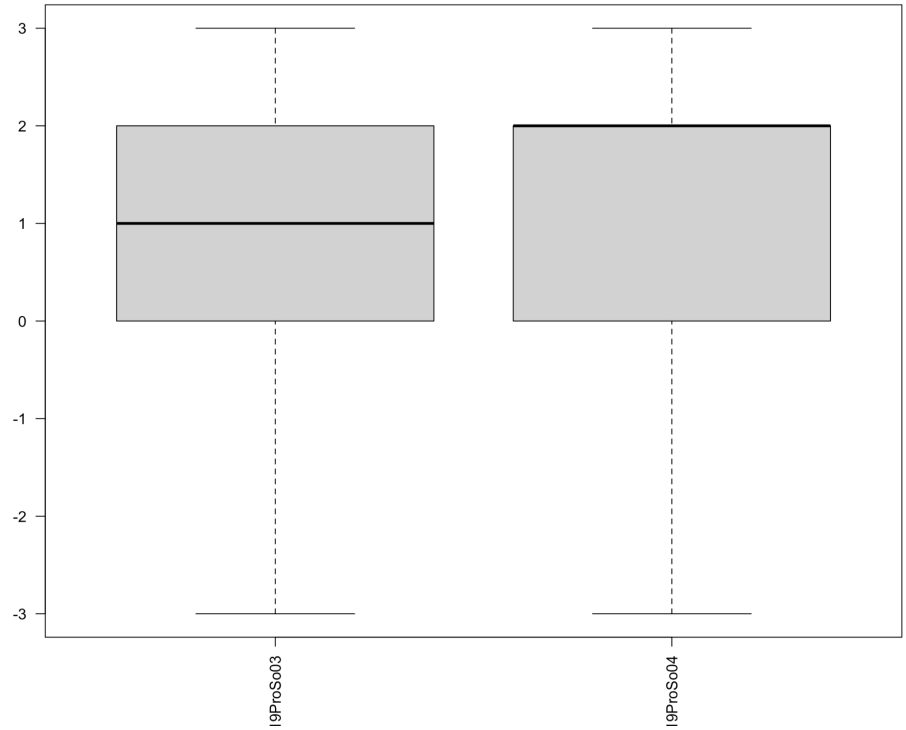


Figure 1.1.4

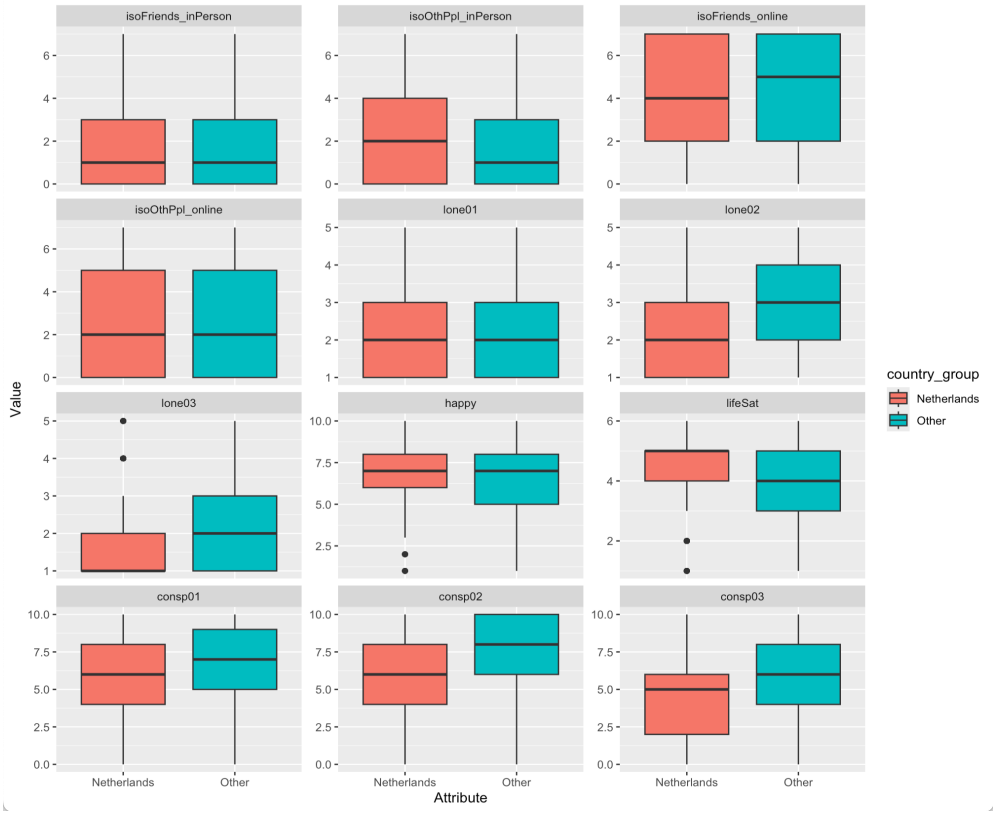


Figure 2.1.1

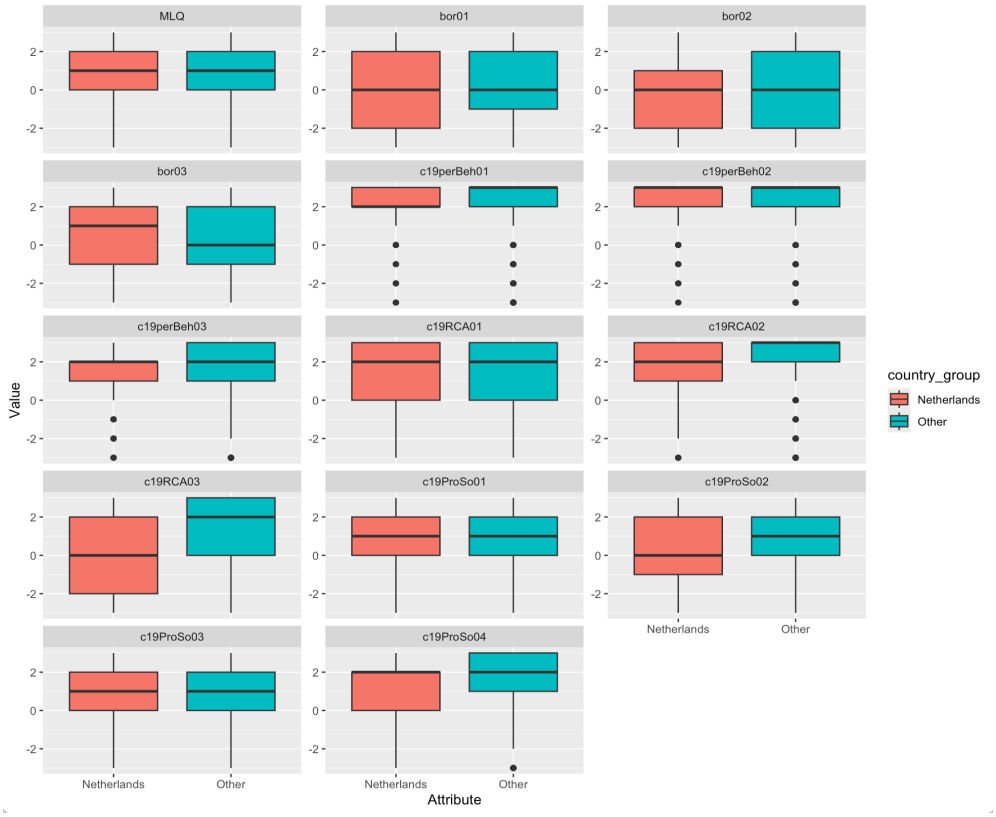


Figure 2.1.2

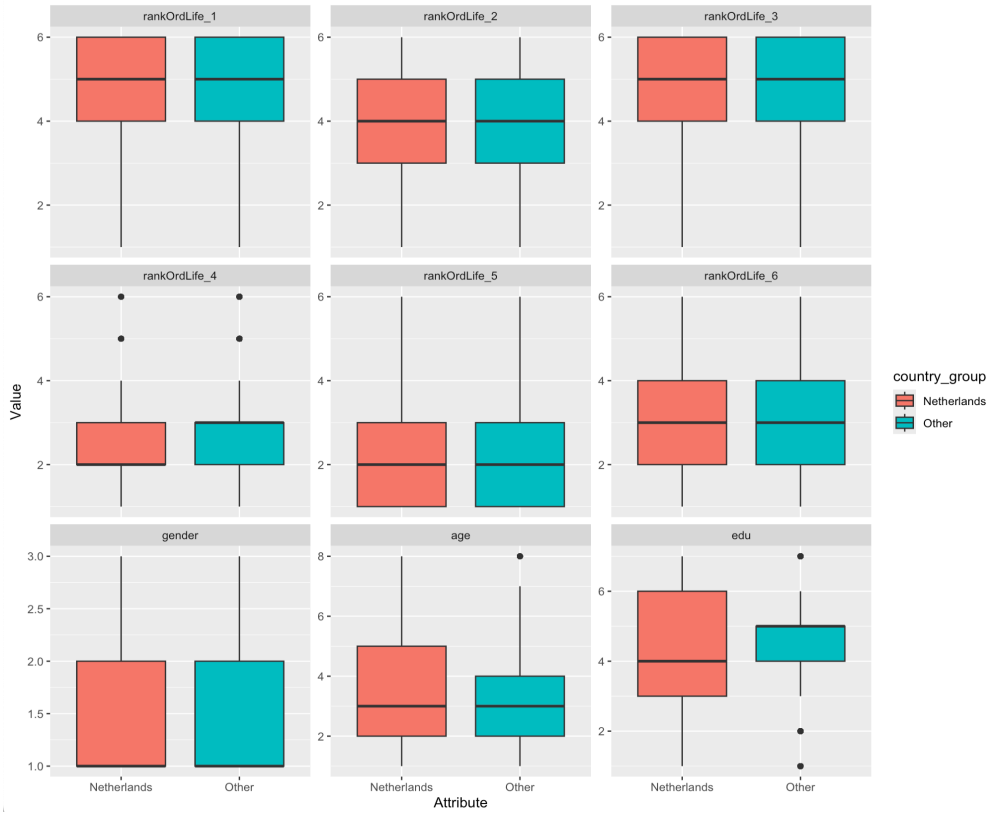


Figure 2.1.3

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(reshape2)
library(cluster)
library(ggrepel)

# reading csv file and create my individual data
rm(list = ls())
set.seed(32909764) # XXXXXXXX = your student ID
cvbase = read.csv("PsyCoronaBaselineExtract.csv")
covid <- cvbase[sample(nrow(cvbase), 40000), ] # 40000 rows

#_____#
# Question 1(a)

# getting the rows and columns of the dataset
dim(covid)

# Viewing the structure of the dataset
str(covid)

# Summary statistics for numerical variables
summary(covid)
```

```
# getting all the variables name
colnames(covid)

# checking the data types for each variable
datatypes <- sapply(covid, class)
table(datatypes)

# boxplot for numerical variable
boxplot(covid[11:26], las=2)
boxplot(covid[33:38], las=2)
boxplot(covid[45:47], las=2)
boxplot(covid[51:52], las=2)

# Get the number of unique countries
length(unique(covid$coded_country))

# Create a data frame with frequencies of each country
country_freq <- as.data.frame(table(covid$coded_country))
country_freq <- country_freq[order(-country_freq$Freq), ]

country_freq

# Create a bar plot with sorted countries and modified axis
ggplot(data = country_freq, aes(x = reorder(Var1, Freq), y = Freq)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Country", y = "Frequency", title = "Distribution of Coded Countries") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#_____________________________________________________________________________#
# Question 1(b)

# Columns to replace NA values with 0
columns_to_change_na_to_0 <- c("employstatus_1", "employstatus_2", "employstatus_3",
                               "employstatus_4", "employstatus_5", "employstatus_6",
                               "employstatus_7", "employstatus_8", "employstatus_9",
                               "employstatus_10", "coronaClose_1", "coronaClose_2",
                               "coronaClose_3", "coronaClose_4", "coronaClose_5",
                               "coronaClose_6")

# Replace NA values with 0 for selected columns
covid <- mutate_at(covid, vars(all_of(columns_to_change_na_to_0)), ~ifelse(is.na(.), 0, .))

# Replace empty strings with NA in the coded_country column
covid <- mutate(covid, coded_country = na_if(coded_country, ""))

# Drop all rows with NA values
covid <- na.omit(covid)

# Define a function to encode categorical variables
encode_categorical <- function(column) {
```

```
  # Define encoding mapping
  encoding <- c("A" = 1, "B" = 2, "C" = 3, "D" = 4, "E" = 5, "F" = 6)
  # Encode the column using the mapping
  encoded <- encoding[column]
  return(encoded)
}

# Apply the encoding function to each column
covid$rankOrdLife_1 <- encode_categorical(covid$rankOrdLife_1)
covid$rankOrdLife_2 <- encode_categorical(covid$rankOrdLife_2)
covid$rankOrdLife_3 <- encode_categorical(covid$rankOrdLife_3)
covid$rankOrdLife_4 <- encode_categorical(covid$rankOrdLife_4)
covid$rankOrdLife_5 <- encode_categorical(covid$rankOrdLife_5)
covid$rankOrdLife_6 <- encode_categorical(covid$rankOrdLife_6)

#_____________________________________________________________________________#
# Question 2(a)

# Filter the dataset for records corresponding to the focus country (Netherlands)
netherlands_data <- covid %>% filter(coded_country == "Netherlands")

# Summary statistics for numerical variables in Focus country
summary(netherlands_data)

# Filter the dataset for records corresponding to other countries
other_countries_data <- covid %>% filter(coded_country != "Netherlands")

# Summary statistics for numerical variables in other countries
summary(other_countries_data)

# created a new dataframe and new column to separate between focus country and other countries
two_groups_df <- covid %>%
  mutate(country_group = ifelse(coded_country == "Netherlands", "Netherlands", "Other"))

df_long <- reshape2::melt(two_groups_df, id.vars = "country_group", measure.vars =
c("isoFriends_inPerson",
                                                    "isoOthPpl_inPerson",
                                                    "isoFriends_online",
                                                    "isoOthPpl_online",
                                                    "lone01", "lone02",
                                                    "lone03", "happy",
                                                    "lifeSat", "consp01",
                                                    "consp02", "consp03"))


# Create side-by-side box plots
ggplot(df_long, aes(x = country_group, y = value, fill = country_group)) +
  geom_boxplot(position = position_dodge(width = 10)) +
  labs(x = "Attribute", y = "Value") +
  facet_wrap(~ variable, scales = "free_y", ncol = 3)
```

```
df_long <- reshape2::melt(two_groups_df, id.vars = "country_group", measure.vars = c("MLQ",
"bor01","bor02",
                                                    "bor03", "c19perBeh01",
                                                    "c19perBeh02", "c19perBeh03",
                                                    "c19RCA01", "c19RCA02",
                                                    "c19RCA03", "c19ProSo01",
                                                    "c19ProSo02", "c19ProSo03",
                                                    "c19ProSo04"))

# Create side-by-side box plots
ggplot(df_long, aes(x = country_group, y = value, fill = country_group)) +
  geom_boxplot(position = position_dodge(width = 10)) +
  labs(x = "Attribute", y = "Value") +
  facet_wrap(~ variable, scales = "free_y", ncol = 3)

df_long <- reshape2::melt(two_groups_df, id.vars = "country_group", measure.vars =
c("rankOrdLife_1", "rankOrdLife_2",
                                                    "rankOrdLife_3", "rankOrdLife_4",
                                                    "rankOrdLife_5", "rankOrdLife_6",
                                                    "gender", "age", "edu"))

# Create side-by-side box plots
ggplot(df_long, aes(x = country_group, y = value, fill = country_group)) +
  geom_boxplot(position = position_dodge(width = 10)) +
  labs(x = "Attribute", y = "Value") +
  facet_wrap(~ variable, scales = "free_y", ncol = 3)

# t-test for isoOthPpl_inPerson
t_test_isoOthPpl_inPerson <- t.test(netherlands_data$isoOthPpl_inPerson,
other_countries_data$isoOthPpl_inPerson)
t_test_isoOthPpl_inPerson

# t-test for consp02
t_test_consp02 <- t.test(netherlands_data$consp02, other_countries_data$consp02)
t_test_consp02

#_____________________________________________________________________________#
# Question 2(b)

# drop coded_country
netherlands_data <- netherlands_data %>% select(-coded_country)

# Perform linear regression for Netherlands
c19ProSo01_lm <- lm(c19ProSo01 ~ ., data = netherlands_data)
c19ProSo02_lm <- lm(c19ProSo02 ~ ., data = netherlands_data)
c19ProSo03_lm <- lm(c19ProSo03 ~ ., data = netherlands_data)
c19ProSo04_lm <- lm(c19ProSo04 ~ ., data = netherlands_data)

# Print summary of the regression model for Netherlands
summary(c19ProSo01_lm)
summary(c19ProSo02_lm)
```

```
summary(c19ProSo03_lm)
summary(c19ProSo04_lm)

# Question 2(c)

# drop coded_country
other_countries_data <- other_countries_data %>% select(-coded_country)

# Perform linear regression for other countries
c19ProSo01_lm_oth <- lm(c19ProSo01 ~ ., data = other_countries_data)
c19ProSo02_lm_oth <- lm(c19ProSo02 ~ ., data = other_countries_data)
c19ProSo03_lm_oth <- lm(c19ProSo03 ~ ., data = other_countries_data)
c19ProSo04_lm_oth <- lm(c19ProSo04 ~ ., data = other_countries_data)

# Print summary of the regression model for other countries
summary(c19ProSo01_lm_oth)
summary(c19ProSo02_lm_oth)
summary(c19ProSo03_lm_oth)
summary(c19ProSo04_lm_oth)

#_____________________________________________________________________________#
# Question 3(a)

# Read indicator datasets
GDP_per_capita_indicator <- read.csv("recent_GDP_capita.csv")
Urban_population_indicator <- read.csv("recent_urban_population.csv")
Rural_population_indicator <- read.csv("recent_rural_population_percent.csv")
birth_rate_indicator <- read.csv("recent_fctb_birth_rate.csv")
death_rate_indicator <- read.csv("recent_fctb_death_rate.csv")

# Merge the indicator datasets by the column "country"
indicator_df <- merge(GDP_per_capita_indicator, Urban_population_indicator, by = "country")
indicator_df <- merge(indicator_df, Rural_population_indicator, by = "country")
indicator_df <- merge(indicator_df, birth_rate_indicator, by = "country")
indicator_df <- merge(indicator_df, death_rate_indicator, by = "country")

# Drop rows with missing values
indicator_df <- na.omit(indicator_df)

# Keep the "country" column as an identifier
countries <- indicator_df$country

# Remove the "country" column before scaling the data
indicator_df <- indicator_df[, -which(names(indicator_df) == "country")]

# Standardize the data
scaled_data <- scale(indicator_df)

# Determine the number of clusters (k)
# elbow method:
wss <- numeric(10)
```

```
for (i in 1:10) {
  wss[i] <- sum(kmeans(scaled_data, centers = i)$withinss)
}
plot(1:10, wss, type = "b", xlab = "Number of clusters", ylab = "Within-cluster sum of squares", main
= "Elbow Method")

# silhouette method:
sil_width <- sapply(2:10, function(k) {
  kmeans_result <- kmeans(scaled_data, centers = k)
  silhouette_width <- silhouette(kmeans_result$cluster, dist(scaled_data))
  mean(silhouette_width[, "sil_width"])
})

# Plot silhouette widths for different numbers of clusters
plot(2:10, sil_width, type = "b", xlab = "Number of clusters", ylab = "Mean Silhouette Width", main =
"Silhouette Method")

# Find the optimal number of clusters with the maximum silhouette width
optimal_clusters <- which.max(sil_width) + 1
cat("Optimal number of clusters based on silhouette method:", optimal_clusters, "\n")

# Perform k-means clustering with the optimal number of clusters
optimal_kmeans <- kmeans(scaled_data, centers = optimal_clusters)

# Add cluster assignments to the indicator dataframe
indicator_df$cluster <- as.factor(optimal_kmeans$cluster)

# Reattach the "countries" vector to the dataframe
indicator_df <- cbind(country = countries, indicator_df)

# Convert cluster column to numeric
indicator_df$cluster <- as.numeric(as.character(indicator_df$cluster))

# Find the cluster assignment of the Netherlands
netherlands_cluster <- as.integer(optimal_kmeans$cluster[indicator_df$country == "Netherlands"])

# Filter dataframe to include only countries in the same cluster as the Netherlands
same_cluster_df <- indicator_df[indicator_df$cluster == netherlands_cluster, ]

# Calculate distances between each country in the same cluster and the Netherlands using Euclidean
distance
country_distances <- apply(same_cluster_df[, -1], 1, function(x) {
  sqrt(sum((x - same_cluster_df[same_cluster_df$country == "Netherlands", -1])^2))
})

# Add distances as a new column to the dataframe
same_cluster_df$distance_to_netherlands <- country_distances

# Sort the dataframe based on distances
sorted_same_cluster_df <- same_cluster_df[order(same_cluster_df$distance_to_netherlands), ]
```

```
# Extract rows 2 to 7 from the sorted dataframe (excluding the Netherlands)
closest_countries <- sorted_same_cluster_df$country[2:7]

# Print the closest countries
closest_countries

# Add cluster assignments to the scaled data
scaled_data_with_clusters <- cbind(scaled_data, cluster = optimal_kmeans$cluster)

# Convert the data to a data frame
scaled_data_with_clusters_df <- as.data.frame(scaled_data_with_clusters)

# Plot the cluster plot
ggplot(scaled_data_with_clusters_df, aes(x = latest.value_NY.GDP.PCAP.CD, y =
latest.value_SP.URB.TOTL.IN.ZS, color = factor(cluster))) +
  geom_point() +
  scale_color_discrete(name = "Cluster") +
  labs(x = "GDP per Capita", y = "Urban Population (% of total)", title = "Cluster Plot") +
  theme_minimal()

#_____________________________________________________________________________#
# Question 3(b)

# Filter the covid dataset with the selected similar countries
similar_countries_data <- covid[covid$coded_country %in% closest_countries, ]

# drop coded_country
similar_countries_data <- similar_countries_data %>% select(-coded_country)

# Linear regression for similar countries
lm_similar_c19ProSo01 <- lm(c19ProSo01 ~ ., data = similar_countries_data)
lm_similar_c19ProSo02 <- lm(c19ProSo02 ~ ., data = similar_countries_data)
lm_similar_c19ProSo03 <- lm(c19ProSo03 ~ ., data = similar_countries_data)
lm_similar_c19ProSo04 <- lm(c19ProSo04 ~ ., data = similar_countries_data)

# Summary of linear regression for similar countries
summary(lm_similar_c19ProSo01)
summary(lm_similar_c19ProSo02)
summary(lm_similar_c19ProSo03)
summary(lm_similar_c19ProSo04)

# Filter the covid dataset without the selected similar countries
other_countries_data <- covid[!covid$coded_country %in% closest_countries, ]

# drop coded_country
other_countries_data <- other_countries_data %>% select(-coded_country)

# Linear regression for other countries
lm_other_c19ProSo01 <- lm(c19ProSo01 ~ ., data = other_countries_data)
lm_other_c19ProSo02 <- lm(c19ProSo02 ~ ., data = other_countries_data)
lm_other_c19ProSo03 <- lm(c19ProSo03 ~ ., data = other_countries_data)
```

```
lm_other_c19ProSo04 <- lm(c19ProSo04 ~ ., data = other_countries_data)

# Summary of linear regression for similar countries
summary(lm_other_c19ProSo01)
summary(lm_other_c19ProSo02)
summary(lm_other_c19ProSo03)
summary(lm_other_c19ProSo04)
```