

task2

data cleaning

```
## Warning in FUN(X[[i]], ...): strings not representable in native encoding will
## be translated to UTF-8
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   country_region = col_character(),
##   a_value = col_double(),
##   b_value = col_double(),
##   c_value = col_double()
## )
```

```
## [1] 27
```

```
## [1] Belarus           Brunei           Cambodia
## [4] China               Denmark          Estonia
## [7] Guatemala           Honduras         Iran
## [10] Jamaica             Japan            Kazakhstan
## [13] Korea, South        Liechtenstein    Norway
## [16] Pakistan            Peru             Qatar
## [19] San Marino          Slovakia          Slovenia
## [22] Sri Lanka           Sweden           Trinidad and Tobago
## [25] Uruguay             Uzbekistan       Venezuela
## 116 Levels: Afghanistan Albania Algeria Andorra Argentina Armenia ... Vietnam
```

```
##      country_region b_value
## 96      Singapore    0.085
```

```
##      country_region b_value
## 106 Trinidad and Tobago    3.857
```

```
## [1] 15
```

```
## [1] Brunei           China            Guatemala
## [4] Honduras         Jamaica          Kazakhstan
## [7] Korea, South     Liechtenstein    San Marino
## [10] Slovakia         Sri Lanka        Trinidad and Tobago
## [13] Uruguay          Uzbekistan       Venezuela
## 116 Levels: Afghanistan Albania Algeria Andorra Argentina Armenia ... Vietnam
```

predict data

```
## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )

## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
```

```

## Last_Update = col_character(),
## Lat = col_double(),
## Long_ = col_double(),
## Confirmed = col_double(),
## Deaths = col_double(),
## Recovered = col_double(),
## Active = col_double(),
## Combined_Key = col_character()
## )
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_character(),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_character(),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )

## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()

```

```
## )
## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_character(),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   FIPS = col_double(),
##   Admin2 = col_character(),
```

```
## Province_State = col_character(),
## Country_Region = col_character(),
## Last_Update = col_character(),
## Lat = col_double(),
## Long_ = col_double(),
## Confirmed = col_double(),
## Deaths = col_double(),
## Recovered = col_double(),
## Active = col_double(),
## Combined_Key = col_character()
## )

## Parsed with column specification:
## cols(
##   FIPS = col_character(),
##   Admin2 = col_character(),
##   Province_State = col_character(),
##   Country_Region = col_character(),
##   Last_Update = col_datetime(format = ""),
##   Lat = col_double(),
##   Long_ = col_double(),
##   Confirmed = col_double(),
##   Deaths = col_double(),
##   Recovered = col_double(),
##   Active = col_double(),
##   Combined_Key = col_character()
## )
```

Afghanistan	3.200053e+03
China	1.211702e+07
Korea_South	9.565317e+05
United_Kingdom	2.690240e+08
US	1.428445e+10
Vietnam	8.978671e+01

country_region	a_value	b_value	c_value
Afghanistan	342	0.202	37
Albania	269	0.173	17
Algeria	723	0.258	30
Andorra	345	0.344	22
Argentina	970	0.315	23
Armenia	514	0.288	23
Australia	4072	0.293	58
Austria	10760	0.275	28
Azerbaijan	365	0.184	30
Bahrain	795	0.118	29
Bangladesh	99	0.244	18
Belarus	102	0.276	19
Belgium	8530	0.254	49
Bolivia	81	0.192	16

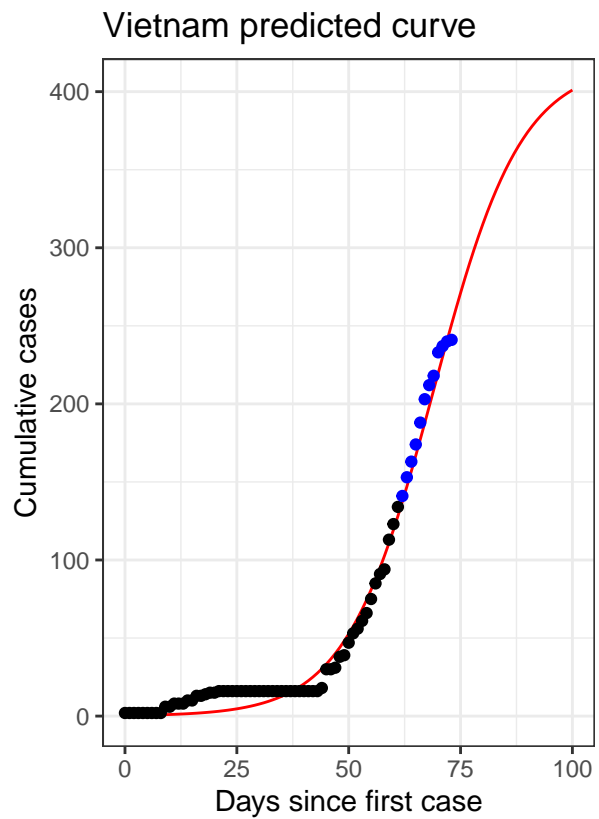
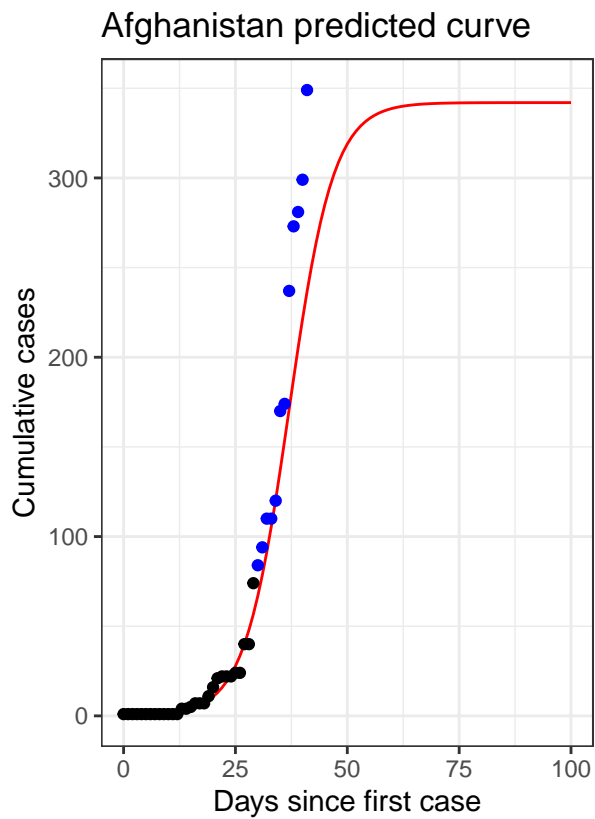
country_region	a_value	b_value	c_value
Bosnia and Herzegovina	352	0.292	19
Brazil	4507	0.380	27
Brunei	98	0.381	7
Bulgaria	459	0.253	16
Burkina Faso	252	0.363	14
Cambodia	168	0.317	56
Canada	5462	0.338	58
Chile	1862	0.318	21
China	78732	0.223	18
Colombia	777	0.335	18
Congo (Kinshasa)	115	0.360	14
Costa Rica	375	0.268	18
Cote d'Ivoire	342	0.857	15
Croatia	958	0.310	29
Cuba	122	0.363	13
Cyprus	272	0.234	15
Denmark	3258	0.170	24
Dominican Republic	640	0.498	23
Ecuador	2180	0.449	23
Egypt	806	0.193	39
Estonia	569	0.235	22
Finland	1570	0.216	55
France	39932	0.148	64
Georgia	151	0.140	29
Germany	65957	0.259	57
Ghana	300	0.332	15
Greece	1499	0.182	27
Guatemala	23	0.589	6
Honduras	32	0.549	8
Hungary	393	0.266	20
Iceland	1311	0.213	25
India	1060	0.253	54
Indonesia	1389	0.266	22
Iran	49441	0.131	33
Iraq	642	0.143	30
Ireland	2673	0.309	24
Israel	4055	0.304	33
Italy	138340	0.183	53
Jamaica	20	0.331	5
Japan	2195	0.094	60
Jordan	326	0.302	21
Kazakhstan	69	0.529	5
Kenya	237	0.320	18
Korea, South	8801	0.284	40
Kuwait	564	0.088	36
Kyrgyzstan	279	0.546	9
Latvia	411	0.270	22
Lebanon	829	0.169	35
Liechtenstein	55	0.500	15
Lithuania	432	0.451	25
Luxembourg	2213	0.354	24
Malaysia	3231	0.222	59

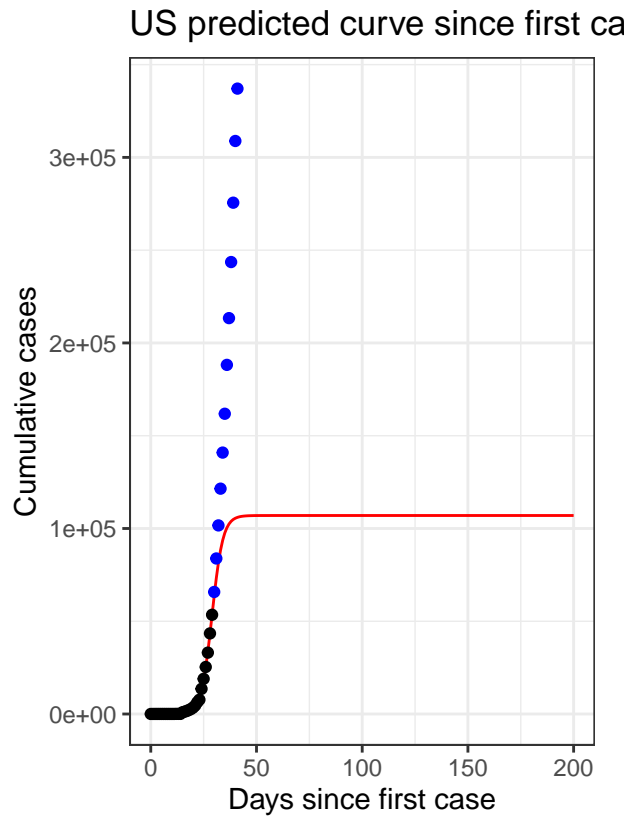
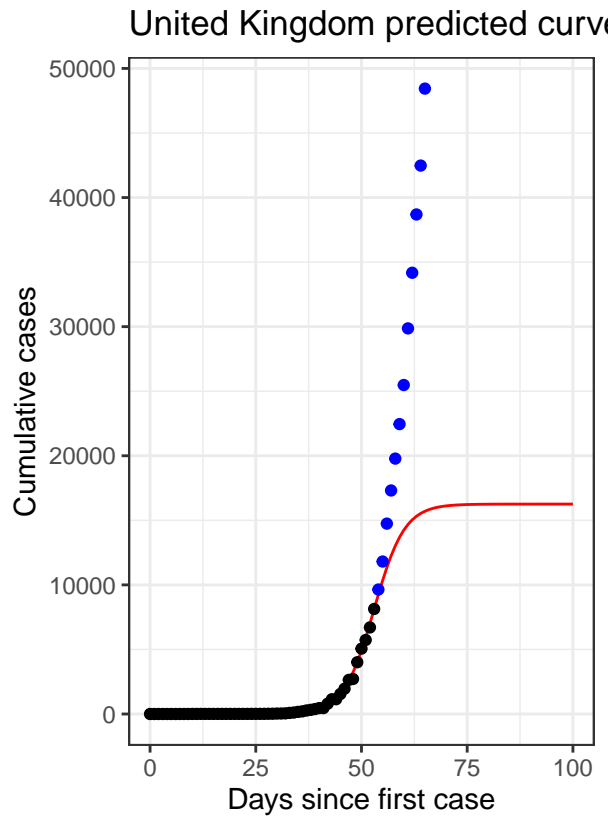
country_region	a_value	b_value	c_value
Malta	242	0.248	17
Martinique	135	0.251	18
Mauritius	115	0.492	7
Mexico	748	0.317	25
Moldova	273	0.285	16
Monaco	60	0.272	25
Montenegro	124	0.507	8
Morocco	357	0.291	22
Netherlands	11170	0.239	26
New Zealand	505	0.420	27
Nigeria	102	0.407	25
North Macedonia	309	0.325	27
Norway	5557	0.175	26
Oman	361	0.125	40
Pakistan	1774	0.326	26
Panama	715	0.321	14
Paraguay	74	0.195	19
Peru	678	0.322	16
Philippines	1091	0.240	54
Poland	1821	0.283	20
Portugal	4741	0.335	22
Qatar	889	0.175	19
Romania	1783	0.256	29
Russia	979	0.291	53
Rwanda	107	0.356	11
San Marino	230	0.191	19
Saudi Arabia	1551	0.288	23
Senegal	357	0.217	27
Serbia	627	0.286	18
Singapore	1262	0.085	67
Slovakia	254	0.332	13
Slovenia	805	0.200	16
South Africa	1303	0.343	20
Spain	79759	0.257	52
Sri Lanka	105	0.459	51
Sweden	4381	0.171	52
Switzerland	19766	0.261	28
Taiwan*	576	0.097	70
Thailand	1634	0.306	62
Trinidad and Tobago	53	3.857	6
Tunisia	419	0.242	24
Turkey	3770	0.537	13
Ukraine	212	0.395	21
United Arab Emirates	652	0.114	62
United Kingdom	16258	0.279	53
Uruguay	184	0.548	6
US	106991	0.389	29
Uzbekistan	50	0.729	4
Venezuela	95	0.426	5
Vietnam	418	0.102	69

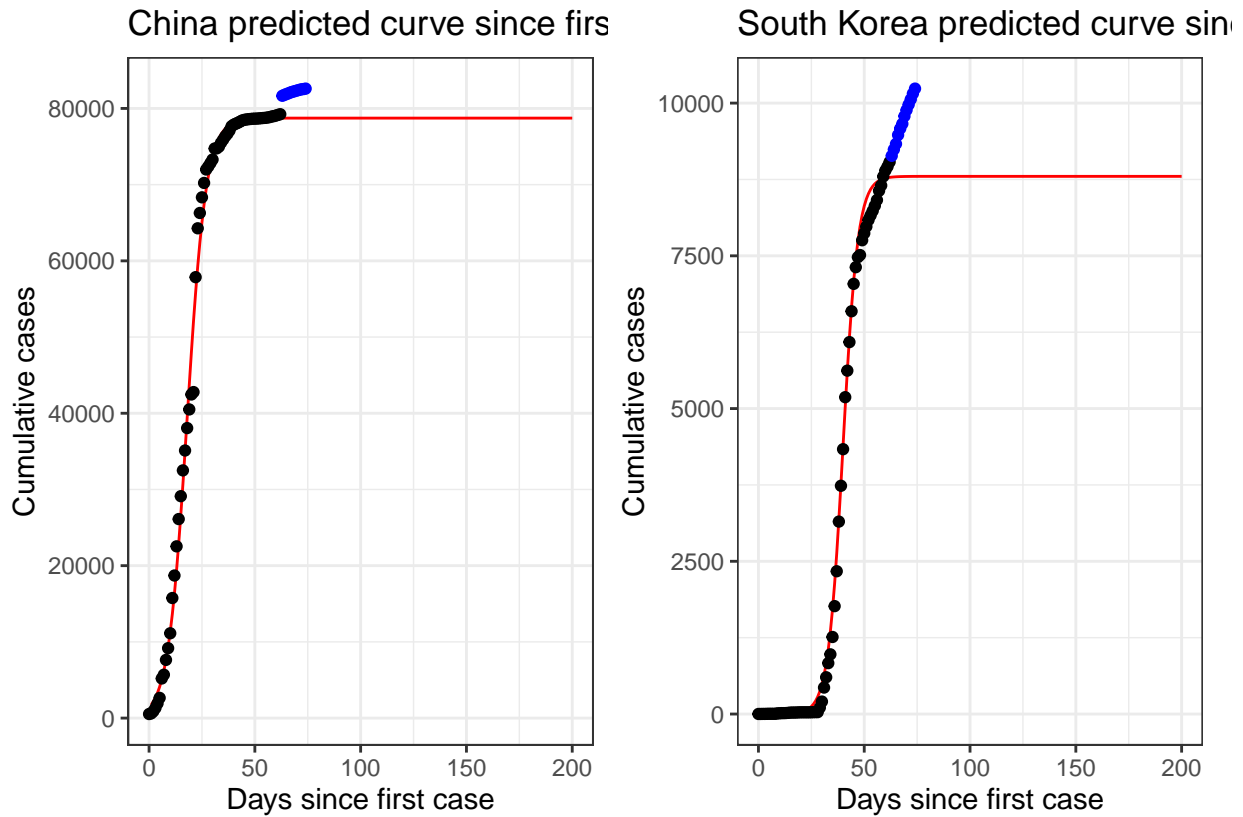
country_region	a_value	b_value	c_value
Afghanistan	342	0.202	37
China	78732	0.223	18
Korea, South	8801	0.284	40
United Kingdom	16258	0.279	53
US	106991	0.389	29
Vietnam	418	0.102	69

Plot some country

```
## Warning: `as.tibble()` is deprecated as of tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```







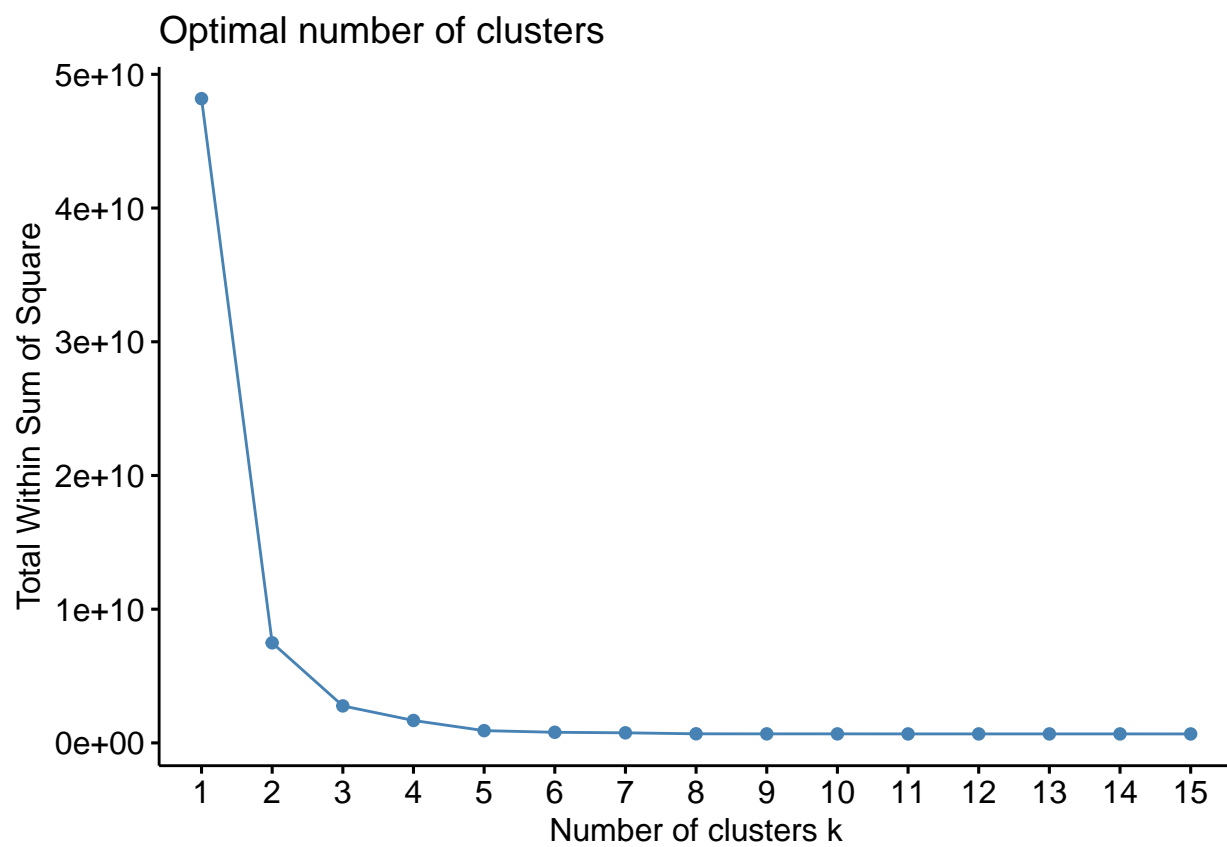
Country	Train_error
Afghanistan	2.080206e+01
China	4.077602e+06
Korea_South	4.471121e+04
United_Kingdom	9.472004e+03
US	1.871744e+05
Vietnam	5.664849e+01

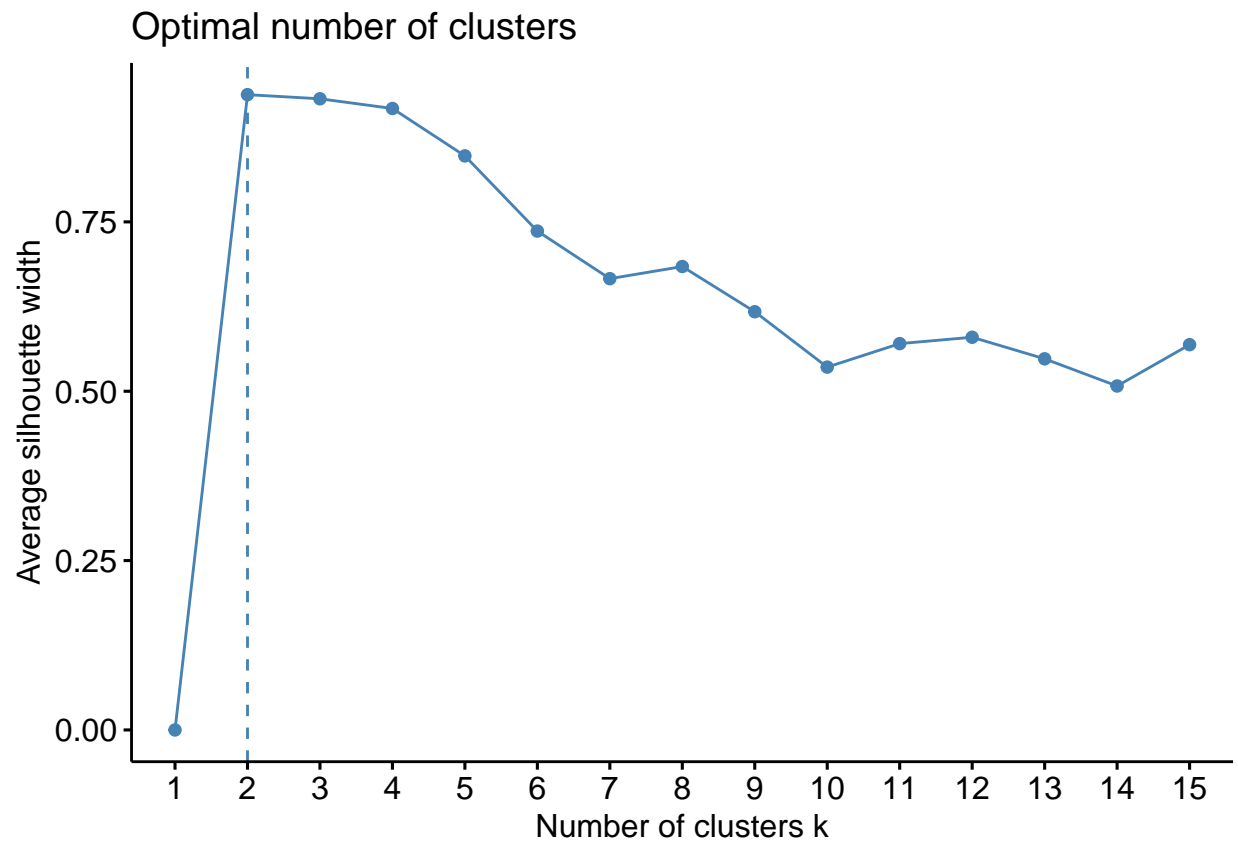
Country	test_error
Afghanistan	3.200053e+03
China	1.211702e+07
Korea_South	9.565317e+05
United_Kingdom	2.690240e+08
US	1.428445e+10
Vietnam	8.978671e+01

Task2

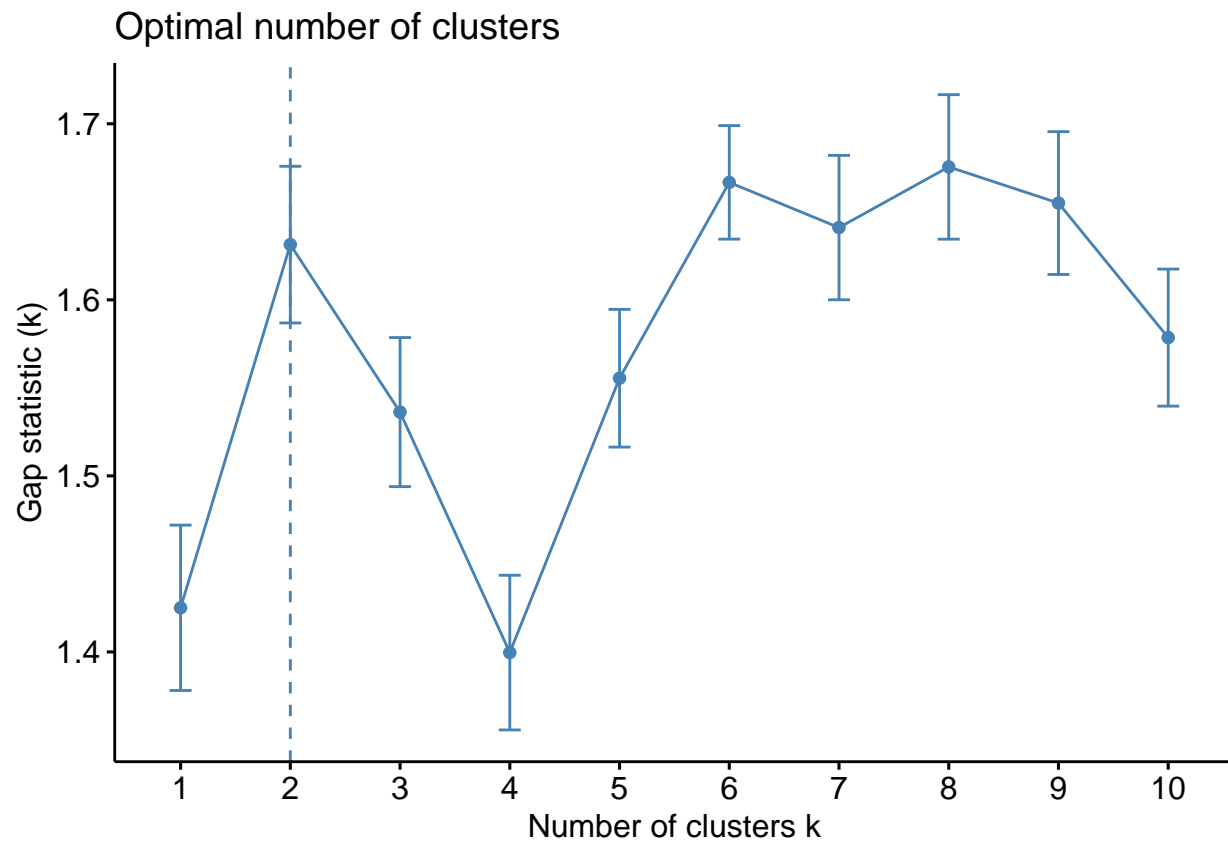
```
## Warning: package 'factoextra' was built under R version 3.6.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```





```
## Warning: package 'cluster' was built under R version 3.6.3
```



```
## [1] 2.826334e+10 1.391574e+01 2.965951e+04
```

```
## [1] 7.485164e+09 1.441893e+01 3.048189e+04
```

```
## $`1`
```

	a_value	b_value	c_value	cluster	country
## 23	78732	0.223	18	1	China
## 39	65957	0.259	57	1	Germany
## 48	49441	0.131	33	1	Iran
## 52	138340	0.183	53	1	Italy
## 100	79759	0.257	52	1	Spain
## 113	106991	0.389	29	1	US

```
##
```

```
## $`2`
```

	a_value	b_value	c_value	cluster	country
## 1	342	0.202	37	2	Afghanistan
## 2	269	0.173	17	2	Albania
## 3	723	0.258	30	2	Algeria
## 4	345	0.344	22	2	Andorra
## 5	970	0.315	23	2	Argentina
## 6	514	0.288	23	2	Armenia
## 7	4072	0.293	58	2	Australia
## 8	10760	0.275	28	2	Austria
## 9	365	0.184	30	2	Azerbaijan
## 10	795	0.118	29	2	Bahrain

## 11	99	0.244	18	2	Bangladesh
## 12	102	0.276	19	2	Belarus
## 13	8530	0.254	49	2	Belgium
## 14	81	0.192	16	2	Bolivia
## 15	352	0.292	19	2	Bosnia and Herzegovina
## 16	4507	0.380	27	2	Brazil
## 17	98	0.381	7	2	Brunei
## 18	459	0.253	16	2	Bulgaria
## 19	252	0.363	14	2	Burkina Faso
## 20	168	0.317	56	2	Cambodia
## 21	5462	0.338	58	2	Canada
## 22	1862	0.318	21	2	Chile
## 24	777	0.335	18	2	Colombia
## 25	115	0.360	14	2	Congo (Kinshasa)
## 26	375	0.268	18	2	Costa Rica
## 27	342	0.857	15	2	Cote d'Ivoire
## 28	958	0.310	29	2	Croatia
## 29	122	0.363	13	2	Cuba
## 30	272	0.234	15	2	Cyprus
## 31	3258	0.170	24	2	Denmark
## 32	640	0.498	23	2	Dominican Republic
## 33	2180	0.449	23	2	Ecuador
## 34	806	0.193	39	2	Egypt
## 35	569	0.235	22	2	Estonia
## 36	1570	0.216	55	2	Finland
## 37	39932	0.148	64	2	France
## 38	151	0.140	29	2	Georgia
## 40	300	0.332	15	2	Ghana
## 41	1499	0.182	27	2	Greece
## 42	23	0.589	6	2	Guatemala
## 43	32	0.549	8	2	Honduras
## 44	393	0.266	20	2	Hungary
## 45	1311	0.213	25	2	Iceland
## 46	1060	0.253	54	2	India
## 47	1389	0.266	22	2	Indonesia
## 49	642	0.143	30	2	Iraq
## 50	2673	0.309	24	2	Ireland
## 51	4055	0.304	33	2	Israel
## 53	20	0.331	5	2	Jamaica
## 54	2195	0.094	60	2	Japan
## 55	326	0.302	21	2	Jordan
## 56	69	0.529	5	2	Kazakhstan
## 57	237	0.320	18	2	Kenya
## 58	8801	0.284	40	2	Korea, South
## 59	564	0.088	36	2	Kuwait
## 60	279	0.546	9	2	Kyrgyzstan
## 61	411	0.270	22	2	Latvia
## 62	829	0.169	35	2	Lebanon
## 63	55	0.500	15	2	Liechtenstein
## 64	432	0.451	25	2	Lithuania
## 65	2213	0.354	24	2	Luxembourg
## 66	3231	0.222	59	2	Malaysia
## 67	242	0.248	17	2	Malta
## 68	135	0.251	18	2	Martinique

## 69	115	0.492	7	2	Mauritius
## 70	748	0.317	25	2	Mexico
## 71	273	0.285	16	2	Moldova
## 72	60	0.272	25	2	Monaco
## 73	124	0.507	8	2	Montenegro
## 74	357	0.291	22	2	Morocco
## 75	11170	0.239	26	2	Netherlands
## 76	505	0.420	27	2	New Zealand
## 77	102	0.407	25	2	Nigeria
## 78	309	0.325	27	2	North Macedonia
## 79	5557	0.175	26	2	Norway
## 80	361	0.125	40	2	Oman
## 81	1774	0.326	26	2	Pakistan
## 82	715	0.321	14	2	Panama
## 83	74	0.195	19	2	Paraguay
## 84	678	0.322	16	2	Peru
## 85	1091	0.240	54	2	Philippines
## 86	1821	0.283	20	2	Poland
## 87	4741	0.335	22	2	Portugal
## 88	889	0.175	19	2	Qatar
## 89	1783	0.256	29	2	Romania
## 90	979	0.291	53	2	Russia
## 91	107	0.356	11	2	Rwanda
## 92	230	0.191	19	2	San Marino
## 93	1551	0.288	23	2	Saudi Arabia
## 94	357	0.217	27	2	Senegal
## 95	627	0.286	18	2	Serbia
## 96	1262	0.085	67	2	Singapore
## 97	254	0.332	13	2	Slovakia
## 98	805	0.200	16	2	Slovenia
## 99	1303	0.343	20	2	South Africa
## 101	105	0.459	51	2	Sri Lanka
## 102	4381	0.171	52	2	Sweden
## 103	19766	0.261	28	2	Switzerland
## 104	576	0.097	70	2	Taiwan*
## 105	1634	0.306	62	2	Thailand
## 106	53	3.857	6	2	Trinidad and Tobago
## 107	419	0.242	24	2	Tunisia
## 108	3770	0.537	13	2	Turkey
## 109	212	0.395	21	2	Ukraine
## 110	652	0.114	62	2	United Arab Emirates
## 111	16258	0.279	53	2	United Kingdom
## 112	184	0.548	6	2	Uruguay
## 114	50	0.729	4	2	Uzbekistan
## 115	95	0.426	5	2	Venezuela
## 116	418	0.102	69	2	Vietnam