

Primele rânduri ale fișierului, prezentat în figura 1.1 conține denumirea coloanelor tabelului ce cuprinde datele statistice ale seriei spațio – temporale. În acest caz primele patru coloane au denumirea: “Province/State”, “Country/Region”, “Lat” și “Long”. Așadar, primele patru coloane conțin informațiile legate de zona geografică, adică, provincia/statul, țara/regiunea, latitudinea și longitudinea, la care se referă seria temporală care este stocată în coloanele următoare. Începând cu coloana a cincea avem stocate datele statistice ale seriilor temporale. Denumirile acestor coloane sunt datele calendaristice, în formatul “luna/ziua/anul”, ale zilelor care alcătuiesc seria temporală.

Figura 1.2 prezintă conținutul aceluiaș fișier prezentat în figura 1.1, de data aceasta formatat sub forma unui tabel.

1	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20
2		Afghanistan	33.0	65.0	0	0	0	0	0	0	0
3		Albania	41.1533	20.1683	0	0	0	0	0	0	0
4		Algeria	28.0339	1.6596	0	0	0	0	0	0	0
5		Andorra	42.5063	1.5218	0	0	0	0	0	0	0
6		Angola	-11.2027	17.8739	0	0	0	0	0	0	0
7		Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0	0
8		Argentina	-38.4161	-63.6167	0	0	0	0	0	0	0
9		Armenia	40.0691	45.0382	0	0	0	0	0	0	0
10	Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0	0	0	0	0
11	New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4	4
12	Northern Territory	Australia	-12.4634	130.8456	0	0	0	0	0	0	0
13	Queensland	Australia	-28.0167	153.4	0	0	0	0	0	0	0
14	South Australia	Australia	-34.9285	138.6007	0	0	0	0	0	0	0
15	Tasmania	Australia	-41.4545	145.9707	0	0	0	0	0	0	0
16	Victoria	Australia	-37.8136	144.9631	0	0	0	0	1	1	1
17	Western Australia	Australia	-31.9505	115.8605	0	0	0	0	0	0	0
18		Austria	47.5162	14.5501	0	0	0	0	0	0	0

Figura 1.2. Seriile spațio – temporale prezentate în formatul unui tabel.

Tabelul din figura 1.2 prezintă datele statistice din seriile spațio – temporale sub forma unui tabel. Primele patru coloane conțin informațiile legate de zonă și localizarea ei prin coordonatele geografice, latitudine și longitudine. Următoarele coloane conțin datele seriilor temporale. Seriile temporale încep cu data de 22 ianuarie 2020 și sunt prezentate în tabelul din figura 1.2 până la data de 28 ianuarie 2020.

Pentru importul datelor statistice din fișierul CSV într-un obiect de tip listă vom folosi funcția **Import** din pachetul informatic Wolfram Mathematica. Funcția Import va primi ca parametri fișierul ce conține datele și formatul său. Fișierul ce trebuie importat va fi selectat folosind funcția **SystemDialogInput** care va fi apelată cu parametrul "FileOpen". Figura 1.3 prezintă liniile de cod pentru apelul funcțiilor SystemDialogInput și Import, precum și conținutul listei în care s-au importat datele în formatul unui tabel.

```
filePath = SystemDialogInput["FileOpen"];
tsConfirmed = Import[filePath, "CSV"];
tsConfirmed // TableView
```

1	2	3	4	5	6	7	8	9	10	11
Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20
	Afghanistan	33.	65.	0	0	0	0	0	0	0
	Albania	41.1533	20.1683	0	0	0	0	0	0	0
	Algeria	28.0339	1.6596	0	0	0	0	0	0	0
	Andorra	42.5063	1.5218	0	0	0	0	0	0	0
	Angola	-11.2027	17.8739	0	0	0	0	0	0	0
	Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0	0
	Argentina	-38.4161	-63.6167	0	0	0	0	0	0	0
	Armenia	40.0691	45.0382	0	0	0	0	0	0	0
Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0	0	0	0	0
New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4	4
Northern Territory	Australia	-12.4634	130.8456	0	0	0	0	0	0	0
Queensland	Australia	-28.0167	153.4	0	0	0	0	0	0	0
South Australia	Australia	-34.9285	138.6007	0	0	0	0	0	0	0
Tasmania	Australia	-41.4545	145.9707	0	0	0	0	0	0	0
Victoria	Australia	-37.8136	144.9631	0	0	0	0	1	1	1
Western Australia	Australia	-31.9505	115.8605	0	0	0	0	0	0	0
	Austria	47.5162	14.5501	0	0	0	0	0	0	0

Figura 1.3. Liniile de cod pentru importul fișierului CSV și redarea datelor importate sub forma de tabel.

Variabila **filePath** va conține calea spre fișierul de tipul CSV, ce conține datele seriilor spațio – temporale. Variabila **tsConfirmed** este de tipul listă și va conține datele importate care sunt prezentate mai jos în formatul **TableView**. Fișierul “time_series_covid19_confirmed_global.csv” utilizat pentru import conține datele statistice ale cazurilor confirmate de **COVID-19**, centralizate de **Johns Hopkins University**, disponibile la adresa https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases?force_layout=desktop.

2) Reprezentarea grafică a datelor statistice geografice

Pentru reprezentarea geografică a datelor statistice importate folosim funcția **GeoGraphics** din pachetul informatic Wolfram Mathematica. Liniile de cod prezentate în figura 2.1 sunt folosite pentru reprezentarea grafică pe hartă la o anumită dată a cazurilor confirmate de **COVID-19** importate în lista **tsConfirmed**.

```
geoDisk =
Table[{GeoStyling[Opacity[0.5]], Blue,
  GeoDisk[{tsConfirmed[[i, 3]], tsConfirmed[[i, 4]],
    (120 000 / zLevel)
    (Piecewise[{{Round[Log[10, tsConfirmed[[i, columnData]]] + 1,
      tsConfirmed[[i, columnData]] > 0}}, 0]}], {i, 2, linesCNo}}];
worldMap = GeoGraphics[{geoDisk},
  GeoRange -> {{latS, latN}, {longV, longE}},
  GeoProjection -> "Equiangular",
  GeoZoomLevel -> Ceiling[zLevel / 2], GeoGridLines -> None,
  Frame -> True, FrameTicks -> None, ImageSize -> {4 × 160, 4 × 80}];
```

Figura 2.1. Liniile de cod folosite pentru reprezentarea geografică a datelor statistice.

În variabila **geoDisk** avem un tabel de discuri localizate pe hartă prin coordonatele **latitudine** și **longitudine** stocate în coloanele 3 respectiv 4 ale listei **tsConfirmed**. Raza discurilor este proporțională pe o scară logaritmică cu numărul cazurilor confirmate, iar variabila **columnData** conține numărul coloanei corespunzătoare datei calendaristice la care se face reprezentarea grafică. Lăsând la o parte prima linie a listei **tsConfirmed** care conține capul de tabel, pentru fiecare linie începând cu linia 2 și până la ultima linie, **linesCNo**, se generează un disc proporțional cu numărul de cazuri din data calendaristică considerată. Reprezentarea grafică se face cu ajutorul funcției **GeoGraphics** care primește ca parametru lista cu discurile conținute în tabelul **geoDisk**. Funcția **GeoGraphics** face o reprezentare a hărții terestre pentru regiunea delimitată de valorile latitudinii și longitudinii setate prin parametrul **GeoRange**, printr-un tip de proiecție în plan setată prin parametrul **GeoProjection** și la o rezoluție setată prin parametrul **GeoZoomLevel**. Reprezentarea grafică returnată de funcția **GeoGraphics** se regăsește în variabila **worldMap**.

Figura 2.2 prezintă reprezentarea grafică obținută cu ajutorul funcției **GeoGraphics** pentru diferite zone geografice și nivele de detaliere. Pentru partea de sus a figurii avem următoarele valori pentru variabilele ce setează parametrul **GeoRange**: **latS**=30, **latN**=66, **longV**=-21 și **longE**=51, iar pentru variabila ce setează parametrul **GeoZoomLevel**: **zLevel**=5. Pentru partea de jos a figurii avem următoarele valori pentru variabilele ce setează parametrul **GeoRange**: **latS**=31, **latN**=49, **longV**=-114 și **longE**=-78, iar pentru variabila ce setează parametrul **GeoZoomLevel**: **zLevel**=10. Reprezentările din figura 2.2 arată doar zonele geografice fără discurile ce descriu datele statistice.

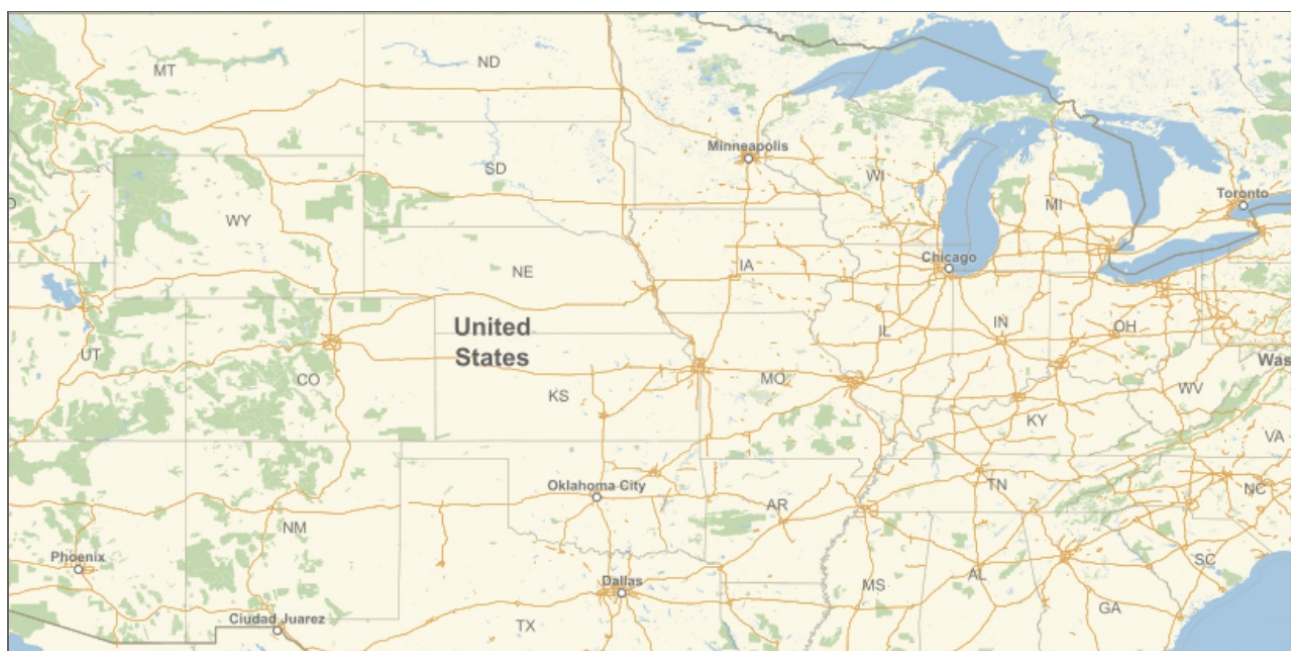


Figura 2.2. Zone geografice reprezentate cu ajutorul funcției **GeoGraphics**, delimitate de diferite valori ale latitudinii și longitudinii, și având diferite nivele de detalieri.

Figura 2.3 prezintă numărul de cazuri de îmbolnăviri **COVID-19** confirmate în zona Europei la două date diferite. În partea de sus a figurii avem reprezentarea datelor statistice la data de 1 martie 2020, iar în partea de jos a figurii avem reprezentarea datelor statistice la data de 31 martie 2020. Numărul cazurilor sunt reprezentate grafic prin discuri de culoare albastră localizate în coordonatele latitudine și longitudine corespunzătoare țărilor care au raportat cazurile. Raza discurilor este proporțională pe o scară logaritmică cu numărul cazurilor raportate de fiecare țară în parte.

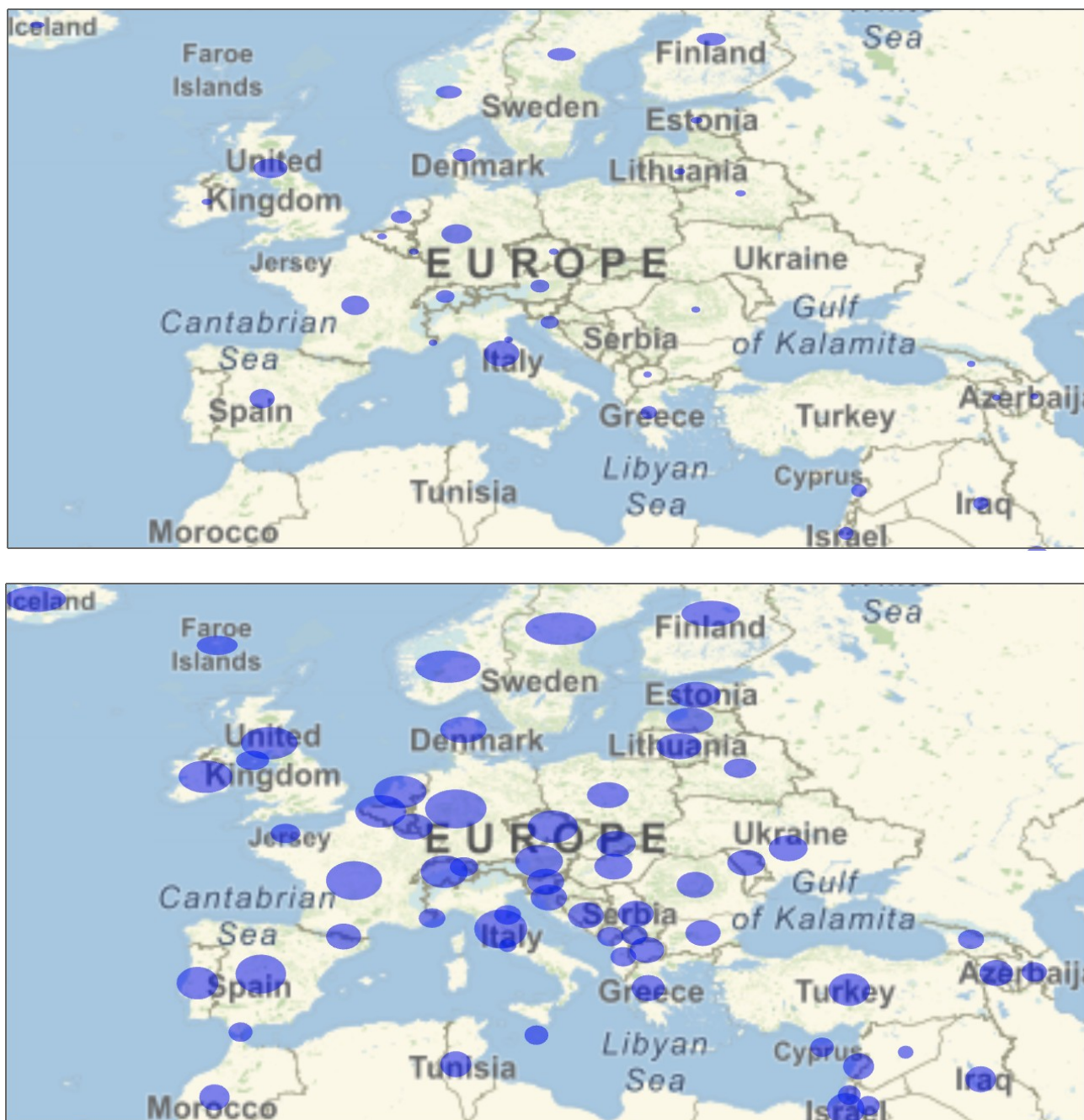


Figura 2.3. Reprezentarea grafică prin discuri a numărului de cazuri de îmbolnăviri COVID-19 în zona Europei la două date diferite: 1 martie 2020 în partea de sus și 31 martie 2020 în partea de jos.

3) Reprezentarea grafică a seriilor statistice temporale

Datele statistice cu privire la cazurile confirmate, recuperate și numărul de decese înregistrate sunt importate din fișierele CSV în listele **tsConfirmed**, **tsRecovered** și respectiv **tsDeaths**. Aceste trei liste au aceeași structură descrisă în prima secțiune pentru lista **tsConfirmed** care conține cazurile confirmate. Pe lângă aceste trei liste creăm o a patra listă **tsActive** care conține seriile temporale pentru cazurile active. Modul de obținere al cazurilor active este evident acestea fiind egale cu numărul de cazuri confirmate din care trebuie să scădem numărul celor recuperați și decedați. Figura 3.1 prezintă liniile de cod folosite pentru a obține seriile temporale corespunzătoare

unei regiuni geografice (țară, stat, provincie, etc.). Aceste serii temporale descriu evoluția în timp a numărului de cazuri confirmate, recuperate, active, precum și numărul celor decedați pe zile într-o zonă geografică.

```
strToDate[data_] := DateObject[{data, {"Month", "/", "Day", "/", "Year"}}];

tsC[[1]] = MapAt[strToDate, tsConfirmed[[1, All]], {5 ;; columnsCNo}];
tsC[[2]] = Select[tsConfirmed, MemberQ[#, state] & MemberQ[#, country] &][[1]];
tsC = Transpose[tsC[[All, 5 ;; columnsCNo]]];
tsR[[1]] = MapAt[strToDate, tsRecovered[[1, All]], {5 ;; columnsRNo}];
tsR[[2]] = Select[tsRecovered, MemberQ[#, state] & MemberQ[#, country] &][[1]];
tsR = Transpose[tsR[[All, 5 ;; columnsRNo]]];
tsD[[1]] = MapAt[strToDate, tsDeaths[[1, All]], {5 ;; columnsDNo}];
tsD[[2]] = Select[tsDeaths, MemberQ[#, state] & MemberQ[#, country] &][[1]];
tsD = Transpose[tsD[[All, 5 ;; columnsDNo]]];
tsA[[1]] = MapAt[strToDate, tsActive[[1, All]], {5 ;; columnsANo}];
tsA[[2]] = Select[tsActive, MemberQ[#, state] & MemberQ[#, country] &][[1]];
tsA = Transpose[tsA[[All, 5 ;; columnsANo]]];

tsPlot = DateListPlot[{tsC, tsR, tsD, tsA}, Joined → False,
  PlotRange → {{strToDate[datePlot], strToDate[dateMax]}, All},
  PlotLabel → StringJoin[{country, " ", state}],
  PlotStyle → {Blue, Darker[Green], Brown, Red},
  PlotLegends →
    PointLegend[{"Confirmed", "Recovered", "Deaths", "Active"},
      LegendMarkers → {{Graphics[Disk[]], 10}}],
  ImageSize → {3 × 180, 3 × 90}];
```

Figura 3.1. Liniile de cod folosite pentru a obține seriile temporale ale cazurilor confirmate, recuperate, active și a numărului de decese pentru o anumită regiune.

După cum am văzut în prima secțiune listele **tsConfirmed**, **tsRecovered**, **tsActive** și **tsDeaths** conțin începând cu coloana a cincea numărul de cazuri confirmate, recuperate, active și respectiv de decese, fiecare coloană corespunzând unei date calendaristice precizate în denumirea coloanei. În prima linie de cod din figura 3.1 definim o funcție numită **strToDate** care convertește o variabilă de tip șir de caractere într-un obiect de tip dată. Cu ajutorul acestei funcții vom putea converti șirurile de caractere din prima linie a listelor **tsConfirmed**, **tsRecovered**, **tsActive** și **tsDeaths** începând cu coloana a cincea până la ultima coloană, dată de variabilele **columnsCNo**, **columnsRNo**, **columnsANo** și respectiv **columnsDNo**, în variabile de tip dată. După aceasta vom extrage valorile numerice ce corespund seriilor temporale reprezentând numărul cazurilor confirmate, recuperate, active și decedate pentru o anumită regiune, specificată prin variabilele **state** și **country**. Valorile conținute în variabilele **state** și **country** se regăsesc în coloana întâia respectiv a doua a listelor **tsConfirmed**, **tsRecovered**, **tsActive** și **tsDeaths**. Pentru o anumită regiune selectată prin valorile variabilelor **state** și **country**, seriile temporale extrase din listele **tsConfirmed**, **tsRecovered**, **tsActive** și **tsDeaths** se regăsesc în listele **tsC**, **tsR**, **tsA** și respectiv **tsD**. Prin aplicarea funcției **Transpose**, listele **tsC**, **tsR**, **tsA** și respectiv **tsD** conțin fiecare două coloane, prima coloană conține datele calendaristice iar a doua valorile corespunzătoare fiecărei date calendaristice. Aceste serii temporale sunt reprezentate cu ajutorul funcției **DateListPlot** care este apelată în finalul secțiunii de cod prezentată în figura 3.1. Variabila **tsPlot** va conține graficul returnat de funcția **DateListPlot** care primește ca prim argument cele patru liste {**tsC**,**tsR**,**tsD**,**tsA**} care trebuie reprezentate grafic. Un alt parametru transmis funcției **DateListPlot** este domeniul temporal al reprezentării grafice, **PlotRange**, având limita inferioară dată de variabila **datePlot**, iar cea superioară dată de variabila

dateMax. De asemenea, titlul graficului este dat de parametrul **PlotLabel** a cărui valoare este atribuită prin concatenarea variabilelor **state** și **country**. Parametrul **PlotLegends** adaugă o legendă graficelor reprezentate.

Figura 3.2 prezintă graficele seriilor temporale ale cazurilor confirmate, recuperate, decedate și active pentru diferite regiuni geografice (țări, state, provincii, etc.) și intervale de timp.

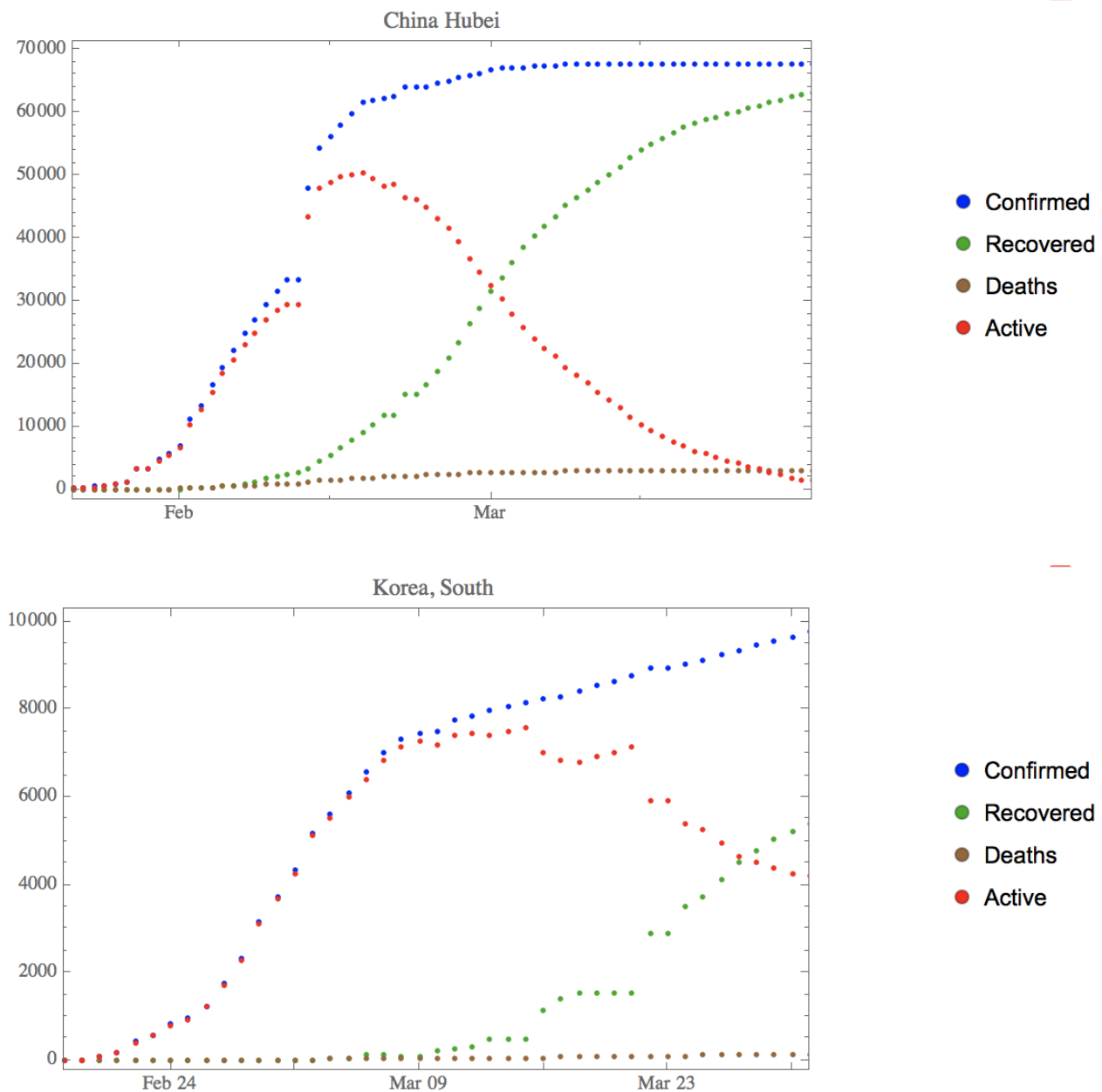


Figura 3.2 Reprezentarea grafică a seriilor statistice temporale reprezentând cazurile de **COVID-19** confirmate, recuperate, decedate și active pentru: provincia Hubei, China, în perioada 22 ianuarie 2020 – 31 martie 2020 (graficul din partea de sus a figurii) și Coreea de Sud, în perioada 18 februarie 2020 – 31 martie 2020 (graficul din partea de jos a figurii).

În partea de sus a figurii 3.2 sunt reprezentate seriile temporale pentru cazurile confirmate, recuperate, decedate și active de **COVID-19** din regiunea Hubei, China, pe perioada 22 ianuarie 2020 – 31 martie 2020. În partea de jos a figurii 3.2 sunt reprezentate seriile temporale pentru cazurile confirmate, recuperate, decedate și active de **COVID-19** Coreea de Sud pe perioada 18 februarie 2020 – 31 martie 2020.

Reprezentările spațio – temporale ale datelor statistice prezentate anterior au fost incluse într-un proiect demonstrativ Wolfram Mathematica. Interfața utilizator precum și reprezentările grafice sunt prezentate în figura 3.3.

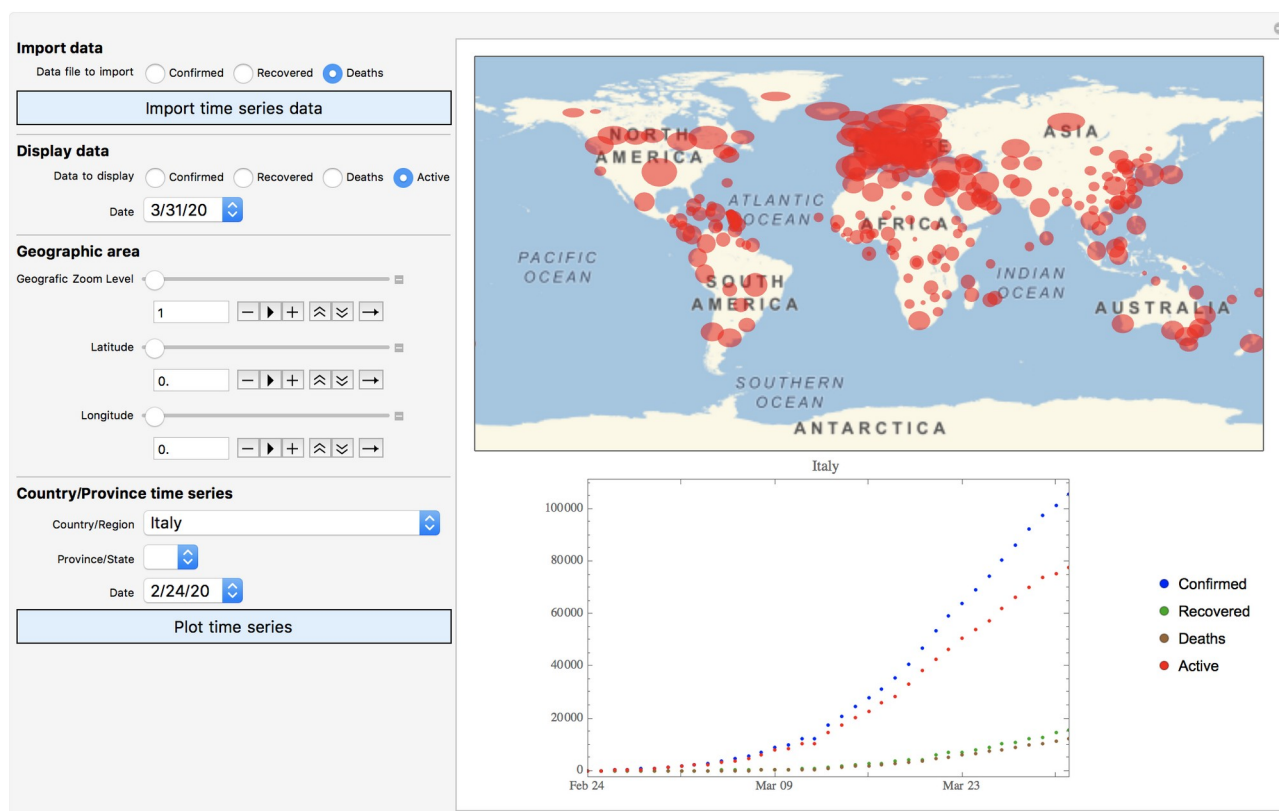


Figura 3.3. Interfața utilizator pentru importarea și reprezentarea grafică a datelor statistice corespunzătoare unor serii spațio – temporale (datele utilizate corespund pandemiei **COVID-19**).

În partea stângă sunt localizate controalele pentru interfața utilizator. Importarea datelor din fișierele CSV se face utilizând controalele din zona care are titlul **“Import data”**. Reprezentarea spațială a datelor la o anumită dată selectată se face utilizând controalele din zona care are titlul **“Display data”**. Modificarea nivelului de detaliere precum și a zonei geografice reprezentate se face utilizând controalele din zona care are titlul **“Geographic area”**. Reprezentarea seriilor temporale pentru o anumită țară și stat/provincie începând cu o anumită dată se face utilizând controalele din zona care are titlul **“Country/Province time series”**. În partea dreaptă sunt localizate reprezentările grafice.

Referințe:

1. Importing and Exporting Files
<https://reference.wolfram.com/language/tutorial/ImportingAndExporting.html#30732>
2. Geographic Data & Entities
<https://reference.wolfram.com/language/guide/GeographicData.html>
3. Time Series Processing
<https://reference.wolfram.com/language/guide/TimeSeries.html>
4. Novel Coronavirus (COVID-19) Cases Data
https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases?force_layout=desktop