

Vizualizarea Datelor - Proiect
Baltariu Ionuț-Alexandru - 1409B
<https://github.com/xeno-john/visualizr>

1.Context [What]

Recent s-au împlinit 2 ani de când virusul SARS COV-2 a început să schimbe iremediabil lumea în care trăim. Încă din prima zi, majoritatea comunităților științifice au început să depună eforturi supraomenești pentru a înțelege virusul și pentru a oferi publicului larg informații concrete și mijloace de prevenție eficiente împotriva acestuia.

Ceea ce este regretabil, însă firesc, este faptul că pentru a înțelege cât mai bine acest virus, oamenii de știință au nevoie de foarte multe date de orice tip despre el. Datele sunt absolut necesare în orice analiză, iar în contextul acestui proiect ele sunt reprezentate de variate statistici corespunzătoare fiecărei zile de la începutul epidemiei în România.

Folosind astfel datele despre COVID-19 (număr de cazuri/teste/infestări/morți/carantinări) ș.a.m.d furnizate de către <https://www.graphs.ro/resources.php>, utilizatorul de rând va putea vizualiza evoluția epidemiei din prima zi până la data curentă, putând să aleagă specific datele pe care dorește să le vadă și punctul temporal reprezentat de acele date.

2.Scop [Why]

Scopul vizualizării anumitor date variază și în funcție de publicul ce realizează acea vizualizare la un anumit moment. Vizualizarea datelor menționate anterior nu este dedicată strict unui anumit public țintă: orice persoană poate observa statisticile despre COVID-19 pentru a se informa despre comportamentul acestei boli sau pentru a fi conștient de efectele acesteia în societatea noastră.

Datele se pot corela, se pot interpreta, se pot contextualiza și lega de factori externi (e.g: relaxarea măsurilor în vara anului 2021 a dus la o creștere semnificativă a cazurilor în toamna aceluiași an, lipsa implicării organelor abilitate în promovarea vaccinării și informării publicului larg despre siguranța acesteia a dus la intrarea în valul IV al epidemiei cu un procent mic de cetățeni vaccinați și eventual la epuizarea resurselor sistemului sanitar, etc).

Presupunând că aceste date sunt raportate într-un mod realist de către instituțiile responsabile, ele pot oferi atât publicului larg cât și specialiștilor o idee fermă despre cum a fost impactată societatea românească în ultimii 2 ani.

3. Detalii tehnice despre proiect [How]

3.1. Obținerea datelor vizate

Pentru obținerea datelor de interes despre pandemia s-a folosit ca sursa <https://www.graphs.ro/resources.php> (Access json here) cu o mențiune: datele nu se preiau la cea mai nouă versiune la fiecare lansare a “aplicației”(paginii HTML) din următorul motiv:

JSON Data Romania New Coronavirus

All the data used for this website is being turned into a JSON feed for you to download and use however you wish as long as you credit the source. We have collected the data from official sources - [The Romanian Ministry of Health](#) and make an effort to keep it accurate and up to date. Please keep in mind that this website is not affiliated with any official Romanian Government Institutions.

Instructions:

- Data is updated daily around 1 PM GMT as per when it is released.
- Where data appears as 'null', there is no official data that we could collect.
- All data is organized by date, starting with the latest update.
- Field names are pretty self explanatory.
- Please do not abuse the system and pull data every minute or you run the risk of getting blocked by our automated systems. Once of twice a day should suffice. Please also cache the data to your servers and do not leech it for live content, this is not an API.

[Access JSON here](#)

[Access Vaccination Data JSON here](#)

Also, a light version with just today's data:

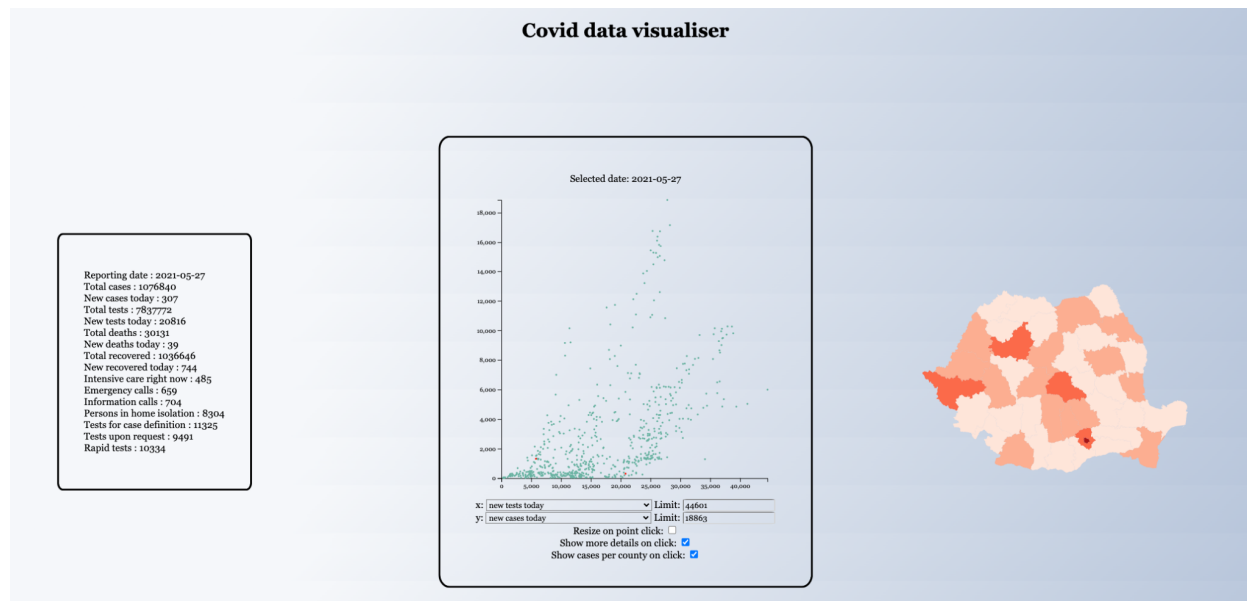
[Access light JSON here](#)

Astfel, ca workaround s-a ales descărcarea json-ului pentru ca acesta sa fie livrat de un server local. Serverul a fost realizat în Python, folosind modulul FastAPI și livrează json-ul unui client care face o cerere HTTP de tip GET la adresa <http://localhost:8000/>. De asemenea, geojson-ul folosit în proiect este disponibil la adresa http://localhost:8000/romania_map și este fix cel din Laboratorul 5 de la disciplina Vizualizarea Datelor.

3.2. Unde se pot observa aceste date

Datele se pot observa vizual în fișierul **app.html**, astfel “aplicația” nucleu a proiectului este o pagină statică HTML. Alegerea a fost făcută în scop demonstrativ, încât mai important era rezultatul decât mijloacele obținerii acestuia.

În realitate(pentru producție) alegerea ar fi fost un framework sau o librărie pentru partea de Front-End, e.g: Angular.js, React.js + Back-End care ar fi obținut datele cele mai de actualitate și le-ar fi preprocesat, etc.



3.3. Sub ce forme se observă datele

După cum se observă în imaginea anterioară, pe pagina statică HTML există 3 informații de interes:

- Chenarul cu informații specifice despre o anumită zi (un punct de pe ScatterPlot)
- Graficul central în care sunt afișate datele de la începutul pandemiei
- Harta României în care se pot observa numărul exact de cazuri dintr-un anumit județ la data selectată (un punct de pe ScatterPlot)

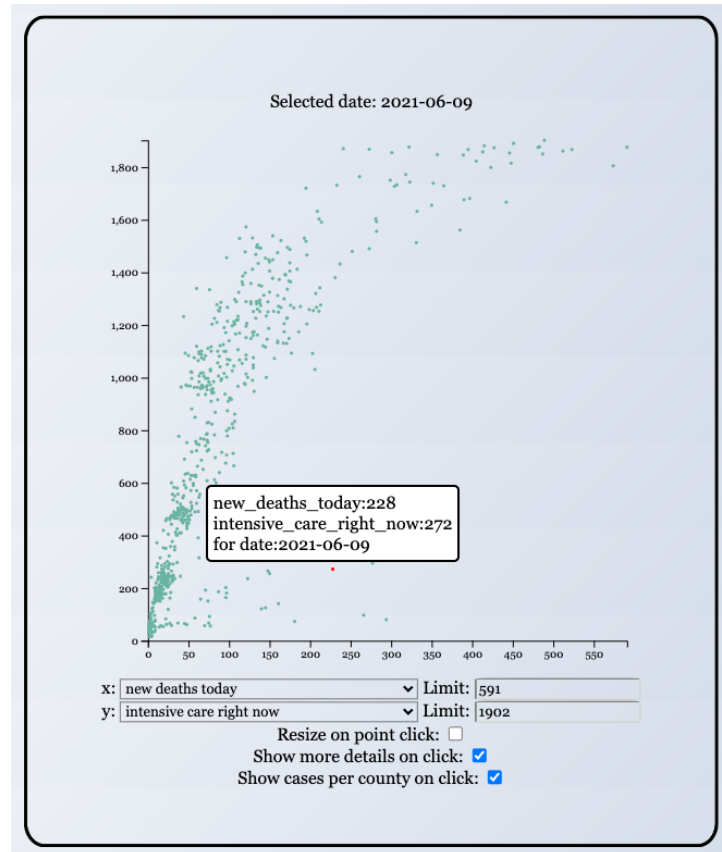
3.3.1. Scatter Plot-ul

Acesta a fost realizat cu gândul de a oferi utilizatorului controlul asupra datelor pe care le observă - detaliile de pe axa x sau y pot fi schimbate în timp real, rezultatul fiind autoscalarea graficului pe ambele axe și redesenarea punctelor.

Dacă utilizatorul dorește să vadă doar acele puncte (1 pct = o zi) care se încadrează într-o limită impusă, indiferent de axă, acesta poate modifica parametrul “Limit” corespunzător fiecărei axe.

De asemenea, dacă se dorește ca un anumit punct să reprezinte extremitatea din dreapta-sus a graficului, după selectarea acestuia graficul se va modifica (opțiunea “*Resize on point click*”).

Ținerea cursorului deasupra oricărui punct de pe grafic rezultă în apariția unei mini-ferestre care va afișa datele exacte (cele selectate) din ziua care îi corespunde acelui punct.



Se observă în imaginea de față faptul că au fost selectate alte metrici pentru axele x și y, iar punctul asupra căruia utilizatorul ține mouse-ul este detaliat în mini-fereastră cu fundalul gri-deschis. Punctul este colorat cu roșu pentru că utilizatorul a și dat click pe acesta (poate fi util atunci când se dorește marcarea anumitor puncte).

3.3.2. Harta colorată după numărul de cazuri a României

Harta apare ca rezultat a unui click pe un anumit punct din graficul menționat anterior, iar rezultatul este o ilustrare grafică a numărului de cazuri din fiecare județ până în punctul temporal selectat (numărul total de cazuri până la o dată aleasă arbitrar).

Culorile folosite au fost indicate de [ColorBrewer](#) și au fost asiguate după numărul de cazuri relativ la maximumul din ziua selectată. Practic au fost create 5 partiții în care se poate situa un anumit județ, urmând ca acestuia să i se dea o culoare în funcție de zona de gravitate în care se încadrează.

```

/* function that helps to fit the given county into a color scheme */
function getColorForNumberOfCases(num_of_cases) {
  const domain = max_county_cases - min_county_cases;
  const partition = domain / 5;
  /*
   the function we use to fit the number of cases
   into a given interval is a 1st grade mathematical function
   bind helps in providing the same constants for all calls
  */
  const f_enhanced = f.bind(null, partition, min_county_cases);
  let color;

  if (num_of_cases > f_enhanced(4)) {
    color = '#a50f15';
  } else if (num_of_cases > f_enhanced(2)) {
    color = '#de2d26';
  } else if (num_of_cases > f_enhanced(1)) {
    color = '#fb6a4a';
  } else if (num_of_cases > f_enhanced(1) / 2) {
    color = '#fcae91';
  } else {
    color = '#fee5d9';
  }

  return color
}

```

f_enhanced reprezintă o funcție de gradul I în care constanta **a** este valoarea unei partiții iar **b** este numărul minim de cazuri din toate județele.

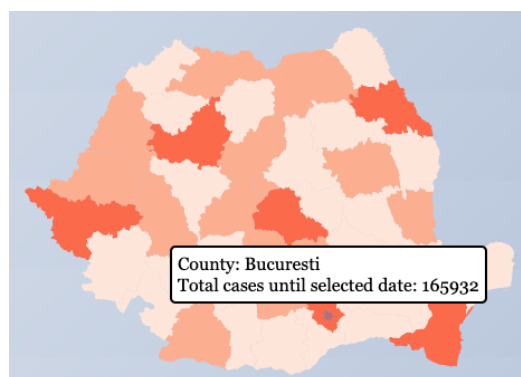
```

/* grade 1 function */
function f(a, b, x) {
  return a * x + b;
}

```

Alegerea **b**-ului este datorată faptului că atunci când se creează domeniul ce urmează a fi partiționat, se scade numărul minim de cazuri din numărul maxim de cazuri corespunzător unui județ. Dacă nu se includea **b**-ul în expresie, majoritatea județelor ar fi căzut în prima categorie de culoare, rezultând într-o hartă care nu dă o idee reală despre situația epidemică la nivel de județ.

Atunci când utilizatorul ține mouse-ul deasupra unui județ, se indică datele exacte de la nivelul acestuia:



3.4. Cum au fost realizate graficele detaliate anterior

Graficele sunt desenate pe pagina HTML sub forma de **SVG** cu ajutorul librăriei [D3](#)(DDD - Data Driven Documents).

Spre exemplu, codul următor rezultă desenarea ScatterPlot-ului:

- Inițial se selectează elementul HTML unde se dorește desenarea graficului și se setează lățimea și înălțimea acestuia apoi se repoziționează graficul pentru a fi vizibil în forma completă.

```
/* append the svg object to the body of the page */
const svg = d3.select("#my_dataviz")
  .append("svg")
  .attr("width", width + margin.left + margin.right)
  .attr("height", height + margin.top + margin.bottom)
  .append("g")
  .attr("transform", `translate(${margin.left}, ${margin.top})`);
```

- Se adaugă axele:

```
/* Add X axis */
const x = d3.scaleLinear()
  .domain([0, Math.max(...x_vals_holder)])
  .range([0, width]);
const xAxis = svg.append("g")
  .attr("transform", `translate(0, ${height})`)
  .call(d3.axisBottom(x));

/* Add Y axis */
const y = d3.scaleLinear()
  .domain([0, Math.max(...y_vals_holder)])
  .range([height, 0]);
const yAxis = svg.append("g")
  .call(d3.axisLeft(y));
```

- Urmând ca apoi să se deseneze punctele corelate cu datele de intrare și să se definească evenimentele care pot avea loc în interacțiunea cu acestea:

```
/* Add dots */
/* x and y attributes are used for the window resize functionality */
svg.append('g')
  .selectAll("dot")
  .data(filtered_data)
  .join("circle")
  .attr("x", function(d) {
    return d[selected_x];
  })
  .attr("cx", function(d) {
    return x(d[selected_x]);
  })
  .attr("y", function(d) {
    return d[selected_y];
  })
  .attr("cy", function(d) {
    return y(d[selected_y]);
  })
  .attr('index', function(d, i) { return +i; })
  .attr("r", 1.5)
  .style("fill", "#69b3a2")
  .on("mouseover.color", mouseover)
  .on("mouseover.tip", tooltip.show)
  .on("mouseleave.color", mouseleave)
  .on("mouseleave.tip;", tooltip.hide)
  .on("click", onclick);

svg.call(tooltip);|
```

Se observă un stil declarativ care permite unui programator să definească în principiu orice fel de structură grafică, să “lege” anumite date de aceasta și să definească interacțiuni cu utilizatorul prin intermediul evenimentelor.

4. Resursele folosite

Graficele realizate cu D3 din acest proiect sunt preluate din variate exemple existente pe internet și adaptate pentru a respecta funcționalitățile propuse (în aceeași manieră cum s-a lucrat la laboratorul disciplinei cu VTK&C++), iar resursele folosite sunt:

- https://www.d3-graph-gallery.com/graph/scatter_basic.html
- https://www.d3-graph-gallery.com/graph/scatter_buttonXlim.html

- <https://bl.ocks.org/curran/966ac9f212a8ceb1ea153fd1ee9f72c4>
- <https://observablehq.com/@weitinglin/how-can-i-skip-an-element-of-an-array-when-i-perform-a-data-data->
- <https://stackoverflow.com/questions/56364915/can-d3-read-geojson-files>
- <https://stackoverflow.com/questions/43646573/d3-get-attributes-from-element/43646752>
- <https://openbase.com/js/d3-v6-tip/documentation>
- <http://bl.ocks.org/lwhitaker3/e8090246a20d9515789b>
- <https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>
- <https://d3js.org/>
- <https://www.d3-graph-gallery.com/index.html>
- Pentru inspirație: laboratoarele de la disciplina VD