Mapping Global Blood Types

Alan Joy, Katherine Roughan

Introduction

For the most part, all human beings can broadly be said to have one of eight blood types[1]. Blood types are a classification of blood based on the presence and absence of certain antibodies and antigens. The presence of these antibodies and antigens are determined by one's alleles, making one's blood type a genetic trait passed down through one's parents [1]. From this, we can assume each country will have a different concentration of each blood type, due to the genetic differences in their population and how they intermingled throughout history.

The ABO blood group system involves two antigens and two antibodies found in human blood, referred to as antigen A, antigen B, antibody A, and antibody B. The presence and absence of these is classified as blood types A, B, AB, and O. The second most significant blood-group system is the Rh system, referring to approximately 50 antigens. However, the most significant of these is the D antigen, because the presence or absence of it is most likely to provoke an immune system response. The presence or absence of this antigen splits the 4 blood groups into 8: A positive, A negative, B positive, B negative, AB positive, AB negative, O positive, and O negative.

The main significance of these blood types lies in the effectiveness and dangers involved in blood transfusion and in birth. If certain people with certain blood types are given blood from people with blood types containing antibodies to the antigens of the recipient, it is likely that the blood transfusion will cause a transfusion reaction, which can destroy red blood cells, cause kidney failure or even death.[2] For pregnant women, the risk lies in the possibility that the fetus carries a blood type different from the mother's. The mother can develop antibodies against the fetus's antigens if the fetus's blood cells pass into the mother's circulatory system, which can cause Rh disease or other hemolytic diseases in the fetus.

Motivation & Audience

As you can probably tell by now, blood types are important, and given that certain countries will have different distributions of each blood type, certain countries will require more of certain blood types. Given this, it would make sense for there to be some sort of mapping for blood types, but upon searching, we could not find any map of this. We felt the need to fill this need ourselves, and so this project was born.

This visualization is designed for usage by organizations like the Red Cross that organize blood drives, so that they might be able to prioritize the distribution of certain blood types to

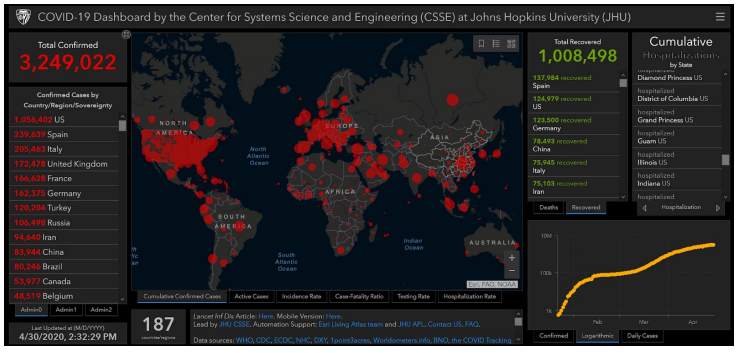
certain countries that may lack people of that blood type. The data for this visualization can be substituted relatively easily, so, if the Red Cross or another relevant organization has their own, more accurate records for blood type distributions by country, they can substitute that data in relatively easily.

We can hypothesise that due to the nature of North America's population being from a wide variety of locations, that North America in particular will have a relatively mixed distribution of blood types. Given that other countries have been around for a very long time, it could be assumed that their blood type distribution would be relatively fixed and not prone to flux, but given that many countries in North America are far far younger, the influx of people from all over the world may still show itself in the blood type distribution of the countries in that area.

Related Work

Due to the long standing importance of geographic data throughout human history, cartographic visualizations have had a long time to evolve both in detail and uses. Computer visualizations are built upon a robust history of advancements in the documentation of cartographic data. With such thorough data available for everyday use both in databases and individual packages and libraries for programming languages, it is now easier than ever to map data in interesting and informative ways. Nöllenburg (2005) explained "the [...] driving force for geovisualization is the need to analyze and explore a dramatically increasing amount of geospatial data[3]." Traditionally, geographic visualizations have had a history in epidemiological uses. Such visualizations could be used to turn a chaotic outbreak into a quantifiable problem that could be planned around. The goal of geographic visualizations is to turn otherwise difficult to understand data on an individually incomprehensible scale into easy-to-understand abstracted regional trends.

One of the most successful recent uses of this integration of cartographic data and large-scale data is Johns Hopkins University's Covid 19 Dashboard[4][5]. This visualization combines real-time updates and analysis with a world map to create a visually clear and meaningful representation of the areas most affected by COVID-19. This visualization features raw numbers, logarithmic and exponential graphs, and geovisualization allowing for the user to approach the data from whatever angle is most useful to



them. Such a multi-faceted approach is something we tried to emulate to some degree in our visualization by including raw data when selecting a region. This however was not as effective as we had hoped and will be discussed further in Feedback.

Design Evolution

The idea began in Figure 1 as a mock-up storyboard of the two categories, individual blood types and largest population. The first goal of the design was to represent blood type presence in a country by means of saturation of the color red. Countries with the greatest percentage population of a given blood type would be a deep red, while countries with the least saturation would be a light pink or white. The second goal of the design was to represent the majority blood type population via a paired color scheme. This color scheme would work well since pairs of light/dark colors can be used to represent positive/negative blood types.

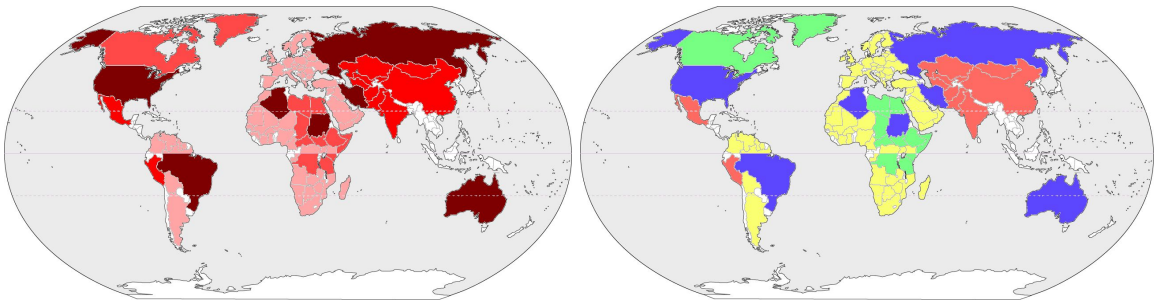


Fig. 1: Storyboard

Following initial storyboarding, we began with the implementation of the map utilizing d3.js and data from NaturalEarthData converted to topojson (Figure 2). During this stage we began experimenting with color and saturation. Several colors, backgrounds colors, stroke colors were tested and sent to individuals for input. Based on their inputs, we chose a white background, gray countries with missing data, and a white border stroke.

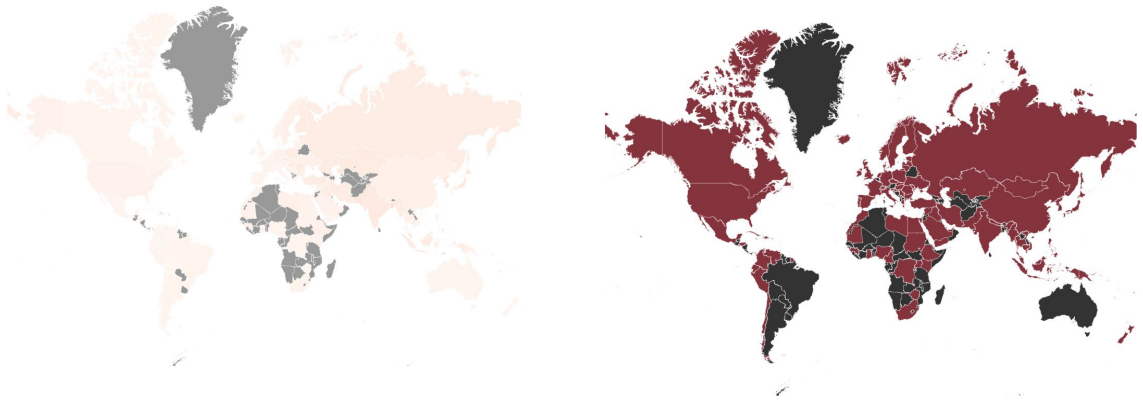


Fig. 2: Prototype

The two main views, percentage population and greatest population can be seen in Figure 3. These represent our final version of the visualization. The saturation scale is adjusted in every view and is now shown in the bottom right of the view. In the greatest population, you can see that there is almost a comedic lack of variation in the colors seen. A paired color scheme is certainly the ideal, however due to the skewed nature of the data it doesn't see particularly successful usage.

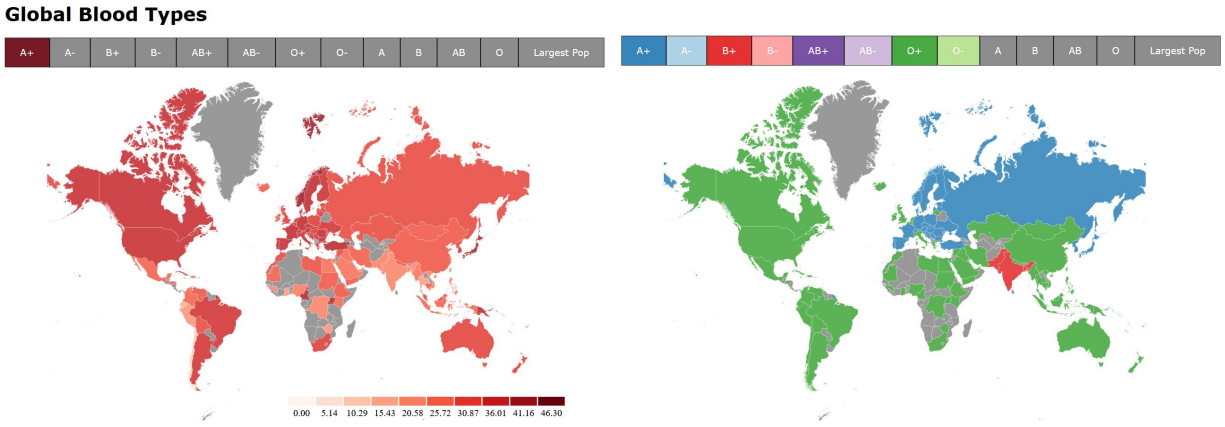


Fig. 3: Final View

Data Acquisition

Due to the sparse and global nature of this data it was rather difficult to track down missing data points. After much contemplation, our data was sourced from the wikipedia article "Blood type distribution by country[6]." This data set has been used in a number of similar visualizations, however we found a few concerning discrepancies in the data set. Notably, the wiki article sources some information from rhesusnegative.net [7]. This website has extremely thorough data *in theory*, but listed only one individual: Lasse Westvang Hougaard and did not

explain the methods by which the data was acquired. Just as well, upon investigating the site it appears to be an almost cultish devotion to the glorification of individuals with Rh-negative blood types. This calls for massive concern in the validity of the data set available on Wikipedia. However, since Wikipedia was the only site we could find the data, we carried on in utilizing what was available to us.

Feedback

Initially, when the idea was first conceived, we received some feedback that we gave some consideration. We were given various D3 toolkit visualization examples that illustrated maps and color scales that we used to build our map and assign colors to the countries. It was suggested that we try to show time-based data, so as to illustrate how blood types change over the years or decades, but ultimately, we lacked the data necessary to make that kind of data worthwhile.

Partway through, we asked some individuals for feedback on the colors used for the visualization. We experimented with other color schemes, such as a black stroke for the outlines of the countries, and a blue background to emulate the ocean, but ultimately decided, based on their feedback, to go with a white background and white stroke for country outlines. This experimentation can be seen in Figure 4. This helped with readability and the overall aesthetics.

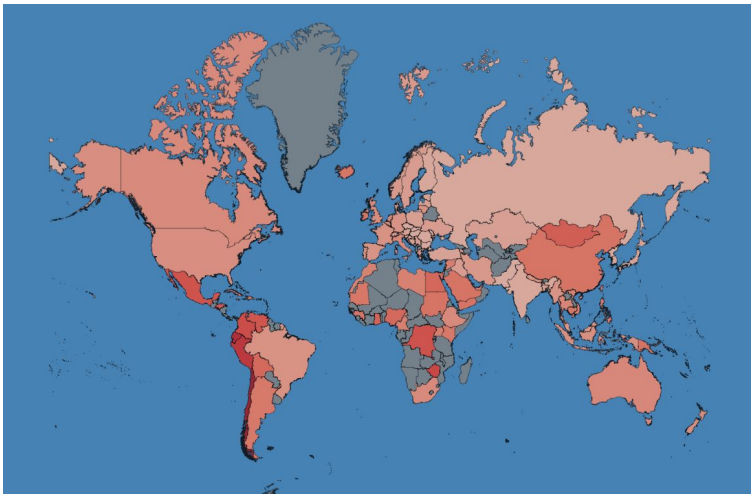


Figure 4. Note the appearance of the smaller, closely grouped countries

After the visualization was built, we received feedback from students and Professor Barbara Cutler. Much of the feedback asked for a feature where users could compare the concentrations of two different blood types for the same country visually. Other ways to visualize the data would have also been useful as well, such as a bubble map. Overall, the feedback indicated that while the map does illustrate the data in a relatively useful way, we could have done more to allow the user to analyze the data by using different types of visualizations.

Results

Overall, the visualization did succeed in visualizing the data in a useful way, and revealed certain trends in the distributions of blood types across the world. If one looks at the Largest Pop view, they can see that the world is primarily made up of blood types A positive and O positive, with some countries having a high concentration of B positive blood types. In particular, you can see that A positive is prevalent in European countries, and that B positive is dominant only in India and the surrounding countries.

However, the hypothesis that North America in particular would be made up of an especially mixed distribution of blood types was incorrect. The United States & Canada are somewhat equally split between A+ and O+ blood types, but other countries in the Americas have blood type distributions similar to those in Asia.

Core Features & Technical Implementation

The visualization is primarily just a map of the world, with each country colored to represent the percentage of individuals in that country with a certain blood type. Buttons are provided for the user to choose between the 8 blood types, as well as buttons that allow the user to ignore Rh alleles when viewing and a button that displays the most concentrated blood type in each country.

The color scale for each blood type is different. We calculated the maximum across all countries for each blood type, and scaled the coloring for each blood type such that the deepest red would always be used for that maximum country. The legend is drawn in the bottom right corner, and enables users to tell exactly what the saturation of a country represents for a given color scale. This allows the difference between countries for a blood type to always be visible regardless of how rare a blood type is. Otherwise, certain blood types, like AB negative, would make the entire map slightly pink, without any noticeable difference in the country with the most AB negative people and the least AB negative people.

The color scale itself is a sequential scale made using d3's scale-chromatic library. The color scheme used for the largest distribution map is a paired color scheme based on colorbrewer. Instead of using the legend from before, the buttons for each blood type are colored. This choice was made in order to save space used up by the legend. Any country that data was missing for was colored a light gray, so as to not contrast too heavily.

The blood type data is read in from a .csv file simply holding the percentages of each blood type for each country, as well as an id for each country. The country data is read in from a json file at the same time, and a d3 queue is used to make sure both datasets are ready when they are processed. The id for each country is used to match the blood type data entry to the

corresponding country geography data entry. The maxes for each blood type are then calculated by going through each country and comparing the blood type values for each country. The blood type data we have in the csv file we used lists blood types including the Rh allele, so to calculate that, we simply added the Rh-positive and Rh-negative percentages and added that to the data saved. The blood types for each country are added to that country in the array of country geography data, and the maxes are added to the top level of the array.

The map itself was implemented using Topojson along with D3's own map-drawing functions. The data for the map itself was pulled from NaturalEarthData, which offers up json data for maps for free. This data was processed using Topojson's processing function, and then colored accordingly. Each country on the map is drawn as a path using D3. The colors of the countries are determined in an update function that is called by buttons at the top of the visualization, as seen in Figure 3. The legend is also created in this function, using the d3-legend library.

Division of Labor

Labor was split between Alan and Katherine based on the step in the process. In the prototype, Alan converted and implemented the NaturalEarthData geographic data in d3. This data was given to Katherine and was in turn converted into csv and used to map available country blood type data from wikipedia data to the country id's in the csv. Alan began processing the data into the saturation view while Katherine tested color schemes and spoke to individuals for input. After the prototype was completed, Katherine made aesthetic/accuracy tweaks to the saturation view and began implementing combination +/- blood type, and largest pop views. During this time, Alan began implementing a tooltip feature as well as a gradient legend. Both members worked on the final aesthetic changes such as the button color and size transitions.

Future Work

In the future, as much of our feedback indicates, it would be good if we added more ways to visualize the data for analysis. In particular, we would add a stacked bar chart that showed the blood type concentrations of each country, so that the user can more easily compare blood types between two countries. In addition, a bubble map would be a good way of showing specific blood distributions for multiple countries at the same time. A feature to compare the blood type distributions for a single country would be useful, since the

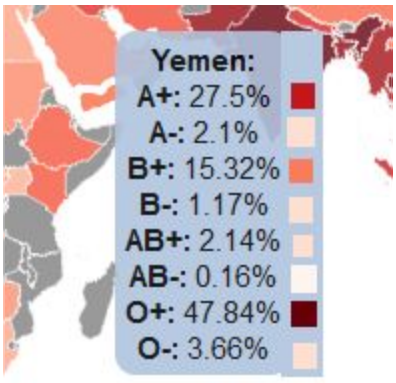


Figure 5. A potential way to compare blood type distributions for a country

differing color scales for the blood types makes it difficult to do that right now. This would most likely be a feature where, once you have selected a blood type, the tooltip for that country shows not only the percentage distribution of blood types for that country, but also the corresponding shade of red for it, with the color scale set to the blood type with the highest percentage for that country. A mockup of this can be seen in Figure 5.

References

1. American Red Cross. "Blood Types." *Explained - A, B, AB and O.* Red Cross Blood Services. www.redcrossblood.org/donate-blood/blood-types.html.

2. BIOTEC Labs (2005). "ABO Blood Group System." *Technical Monograph No. 2.* March 2005. https://web.archive.org/web/20070206095952/http://www.biotec.com/pdf/Technical%20Monograph%20No.%202%20-%20ABO%20system%20and%20subgroups.pdf

3. Martin Nöllenburg (2005). "Geographic Visualization." *Human-Centered Visualization Environments, GI-Dagstuhl Research Seminar.* Dagstuhl Castle, Germany, March 5-8, 2006. https://www.researchgate.net/publication/221025187_Geographic_Visualization

4. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19. https://doi.org/10.1016/S1473-3099(20)30120-1.

5. Center for Systems Science and Engineering at Johns Hopkins University (2020). "COVID-19 Dashboard." https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

6. Blood type distribution by country. (2020, April 26). Retrieved from https://en.wikipedia.org/wiki/Blood_type_distribution_by_country

7. Blood Type Frequencies by Country including the Rh Factor (n.d.). Retrieved from http://www.rhesusnegative.net/themission/bloodtypefrequencies/

Acknowledgements

Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.