École Polytechnique de Montréal

Department of software and computer engineering

INF8808: Data visualization

Groupe 01 (B1)

Mockups

Summer 2020

Sophie Baillargeon-Laporte (1905232)
Philippe Ballandras (1895038)
Meziane Chebili (1899977)

June 6, 2020

1. Overall Context

The COVID-19 pandemic is probably everyone's biggest concern of the year 2020. Its name itself attracts a lot of attention. It's all every news outlet talks about. A good starting point to understand this pandemic is to turn to data by creating insightful visualizations. From there, knowledge and understanding will hopefully arise. Data from governments, health organisations and others concerning various elements of the pandemic are available for us to plot, graph and illustrate the various hidden relationships between this pandemic mechanics and our communities.

2. Goal/Objective

The goal of this project is to provide a view of the evolution of the pandemic through multiple types of data from diverse sources. The objective is to help, through an appealing and efficient data visualisation, the user view how the pandemic affected Canada and entice them to seek more information about the pandemic. The goal is also to use the most data and variables to be able to present the most precise view of the situation.

3. Dataset

The dataset of interest is transmitted by the client, *Le Devoir* independent newspaper. It is composed of data related to the COVID-19 pandemic in Canada, Quebec and Montreal gathered by affiliated journalists from various Health Agencies. Data is available for Canadian provinces, Quebec's regions and Montreal urban agglomeration's regions.

The access to the dataset is made possible through a shared document exposing the data of 10 different spreadsheets. The spreadsheets are not annotated and lack official titles, but the significance of the data is confirmed with the client. Sources are also not indicated, but spreadsheets are classified by source.

The following tables list the different data the project is to exploit and present their description, typical values or range as well as their data type.

The tables are titled with a uniformized version of their original tab names from the document from clearer reference.

Table 3.1: Canada

Variable	Description	Typical values/range	Variable Data Type
Number of Cases of COVID-19	Numerical value of the number of cases of COVID-19	[0, 26594]	Numerical (Ratio)
Number of Repatriated Canadians	Numerical value of the number of repatriated Canadians	[0, 13]	Numerical (Ratio)
Canadian Provinces and Territories	Abbreviations of the Canadian provinces' and territories' name	NL, PE, NS, NB, QC, ON, MB, SK, AB, BC, YT, NT, NU	Nominal
Dates	Calendar dates	[Jan 26th 2020, Apr 30th 2020]	Ordinal

Table 3.2: Canada - Details

Variable	Description	Typical values/range	Variable Data Type
Number of Deaths from COVID-19	Numerical value of the number of death from COVID-19	[0, 1761]	Numerical (Ratio)
Number of Recovered Patients of COVID-19	Numerical value of the number of people who recovered from COVID-19	[0, 8525]	Numerical (Ratio)
Number of Repatriated Canadians	Numerical value of the number of repatriated Canadians	Number of repatriated Canadians: [0, 13]	Numerical (Ratio)
Number of Confirmed Cases of COVID-19	Numerical value of confirmed cases of COVID-19	[0, 26 594]	Numerical (Ratio)
Canadian Provinces and Territories	Abbreviations of the Canadian provinces' and territories' name	NL, PE, NS, NB, QC, ON, MB, SK, AB, BC, YT, NT, NU	Nominal

Table 3.3: Quebec

Variable	Description	Typical values/range	Variable Data Type
Number of Cases of COVID-19	Numerical value of the number of cases of COVID-19	[0, 26594]	Numerical (Ratio)
Code	Quebec regions' code	[1, 18]	Nominal
Regions of Quebec	Full names of the regions of Quebec	Bas Saint-Laurent Saguenay-Lac-Saint-Jean Capitale-Nationale Mauricie-Centre-du-Québec Estrie Montréal Outaouais Abitibi-Témiscamingue Côte-Nord Nord-du-Québec Gaspésie-Îles-de-la-Madeleine Chaudière-Appalaches Laval Lanaudière Laurentides Montérégie Nunavik Terres-Cries-de-la-Baie-James Hors Québec À déterminer	Nominal
Dates	Calendar dates	[Feb 28th 2020, Apr 290th 2020]	Ordinal

Table 3.4: Quebec - Hospitalizations

Variable	Description	Typical values/range	Variable Data Type
Number of Hospitalizations	Numerical value of the number of cases of COVID-19	[0,1648]	Numerical (Ratio)
Number of Intensive Care Cases	Number of Intensive Care Cases	[0,222]	Numerical (Ratio)
Number of cases among health workers	Number of cases among health workers	[0, 46]	Numerical (Ratio)
Dates	Calendar dates	[Jan 26th 2020, Apr 29th 2020]	Ordinal

Table 3.5: Montreal

Variable	Description	Typical values/range	Variable Data Type
Number of Cases of COVID-19	Numerical value of the number of cases of COVID-19	[0, 1153]	Numerical (Ratio)
Codes of Territories of the Urban Agglomeration of Montreal	Codes of Territories of the Urban Agglomeration	2, 4, 5, 6, 7, 8, 9, 10, 11, 16, 22, 28, 29, 38, 39, 43, 44, 45, 46, 47, 50, 51, 52, 54, 56, 57, 59, 60, 61, 62, 63, 64, 65, 66	Nominal
Type of the Territories of the Urban Agglomeration of Montreal	Type describing whether the territory of the urban agglomeration of Montreal is a borough or a city	a (for borough) v (for city)	Nominal
Territories of the Urban Agglomeration	Full names of the regions of the urban	Ahuntsic-Cartierville Anjou Baie-D'Urfé	Nominal

of Montreal	agglomeration of Montreal	Beaconsfield Côte-des-Neiges-Notre-Dame-de-Grâce Côte-Saint-Luc Dollard-des-Ormeaux Dorval Hampstead Kirkland Lachine LaSalle L'Île-Bizard-Sainte-Geneviève L'Île-Dorval Mercier-Hochelaga-Maisonneuve Montréal-Est Montréal-Nord Montréal-Ouest Mont-Royal Outremont Pierrefonds-Roxboro Plateau-Mont-Royal Pointe-Claire Rivière-des-Prairies-Pointe-aux-Trembles Rosemont-La Petite Patrie Sainte-Anne-de-Bellevue Saint-Laurent Saint-Léonard Senneville Sud-Ouest Verdun Ville-Marie Villeray-Saint-Michel-Parc-Extension Westmount Territoire à confirmer	
Dates	Calendar dates	[Mar 28th 2020, Apr 27th 2020]	Ordinal

Table 3.6: Canada - Population

Variable	Description	Typical values/range	Variable Data Type
Canadian Provinces and Territories	Abbreviations of the Canadian provinces' and territories' name	NL, PE, NS, NB, QC, ON, MB, SK, AB, BC, YT, NT, NU	Nominal
Population	Numerical value of individuals living in Canada	[39097, 14711827]	Numerical (ratio)

Table 3.7: Canada - Hospitalizations, Intensive Care, Deaths

Variable	Description	Typical values/range	Variable Data Type
Dates	Calendar dates	[Mar 18th 2020, 11 avril]	Ordinal
Canadian Provinces and Territories	Abbreviations of the Canadian provinces' and territories' name	NL, PE, NS, NB, QC, ON, MB, SK, AB, BC, YT, NT, NU	Nominal
Hospitalisations	Numerical value of the number of hospitalisations	[0, 778]	Numerical (ratio)
Intensive Care	Numerical value of the number of people in intensive care	[0, 264]	Numerical (ratio)
Deaths	Numerical value of the number of COVID-19 deaths	[0, 289]	Numerical (ratio)

Table 3.8: Quebec - Population

Variable	Description	Typical values/range	Variable Data Type
Code	Standard region code	[0, 18]	Nominal
Quebec regions	full name of the Quebec regions	Bas-Saint-Laurent Saguenay–Lac-Saint-Jean Capitale-Nationale Mauricie–Centre-du-Québec Estrie Montréal Outaouais Abitibi-Témiscamingue Côte-Nord Nord-du-Québec Gaspésie–Îles-de-la-Madeleine Chaudière-Appalaches Laval Lanaudière Laurentides Montérégie Nunavik Terres-Cries-de-la-Baie-James	Nominal
Population	Numerical value of individuals living in regions of Quebec	[13798 ; 2065694]	Numerical (Ratio)

Table 3.9: Montreal - Population

Variable	Description	Typical values/range	Variable Data Type
Codes of Territories of the Urban Agglomeration of Montreal	Codes of Territories of the Urban Agglomeration	2, 4, 5, 6, 7, 8, 9, 10, 11, 16, 22, 28, 29, 38, 39, 43, 44, 45, 46, 47, 50, 51, 52, 54, 56, 57, 59, 60, 61, 62, 63, 64, 65, 66	Nominal

Territories of the Urban Agglomeration of Montreal	Full names of the regions of the urban agglomeration of Montreal	Ahuntsic-Cartierville Anjou Baie-D'Urfé Beaconsfield Côte-des-Neiges-Notre-Dame-de-Grâce Côte-Saint-Luc Dollard-des-Ormeaux Dorval Hampstead Kirkland Lachine LaSalle L'Île-Bizard-Sainte-Geneviève L'Île-Dorval Mercier-Hochelaga-Maisonneuve Montréal-Est Montréal-Nord Montréal-Ouest Mont-Royal Outremont Pierrefonds-Roxboro Plateau-Mont-Royal Pointe-Claire Rivière-des-Prairies-Pointe-aux-Trembles Rosemont-La Petite Patrie Sainte-Anne-de-Bellevue Saint-Laurent Saint-Léonard Senneville Sud-Ouest Verdun Ville-Marie Villeray-Saint-Michel-Parc-Extension Westmount Territoire à confirmer	Nominal
Population	Numerical value of individuals living in Canada	[5, 166520]	Numerical (ratio)

Table 3.10: Miscellaneous Data on COVID-19 Infected

Variable	Description	Typical values/range	Variable Data Type
Identification Number of the COVID-19 Case	Identification Number of the COVID-19 Case	[1, 3093]	Ordinal
Reference Period	Reference Period	2020	Nominal
COVID-19 Episode Month	Date of the Month of the COVID-19 episode	1;2;3	
COVID-19 Episode Day	Date of the day of the COVID-19 episode	[01 ; 27]	Ordinal
COVID-19 Episode Date	Full date of the COVID-19 episode	Date in (YYYY-MM-DD) format [Jan 15th 2020, Mar 27th 2020]	Ordinal
Gender of the COVID-19 Infected	Gender of the COVID-19 infected	female male	Nominal
Age Group of the COVID-19 Infected	Age group of the COVID-19 infected in years	20-39 40-49 50-59 60-69 70-79 80+ undeclared	Numerical (Interval)
Transmission	Transmission type of the COVID-19	community transmission travel exposition	Nominal
Hospitalization of the COVID-19 Infected	Whether the COVID-19 infected required hospitalization	yes no unknown undeclared	Nominal
Intensive Care Unit	Whether the COVID-19 infected required	yes no unknown	Nominal

	intensive care		
Status of the Hospitalization Case	Status of the hospitalization case	1: Hospitalized and in intensive care unit 2: Hospitalized, but not in intensive care unit 3: Not hospitalized 9: Not Stated/Unknown	Nominal

4. Target users

In the project context, the client from *Le Devoir* explicitly described the specific target users as numerical readers of *Le Devoir* who are professional adults with a post-secondary education in their thirties from Quebec and have experience in reading through infographics and data visualizations. They also have access to a device allowing digital media viewing such as a computer or a tablet. Their main interest would be to visualize data related to COVID-19 evolution in the Canada and Quebec regions if they deem the visualization interesting at first sight . Hence, they correspond to the profile of the client, Valérie Duhaime.

5. Target Questions

The current section presents the target tasks considered for the data visualization to produce in the shape of target questions. The target questions are distributed into 3 main angles : comparison between geographical regions, comparison between social groups and comparison between COVID-19 infected. A priority is assigned for each question according to its supposed relevance to the target users, a score of 5 being the highest priority and 1 being the lowest.

Angles :

- Comparison between social groups
- Comparison between COVID-19 infected
- Comparison between geographical regions

Table 5.1: Questions of interest for the users evaluated by subjective priority

Questions	Priority (/5)	Vis
1. Who was the age group most affected by COVID-19 in its early stage?	4	V4
2. Who is currently the age group most affected by COVID-19?	4	V4
3. Which gender has more complications in COVID-19 cases?	3	V5
4. How health workers are affected by the virus?	2	
5. Are men more infected than women?	1	V5
6. What proportion of hospitalization required intensive care in Quebec?	3	
7. What proportion of cases required intensive care in Canada?	3	V5
8. What proportion of the infections is transmitted by travel exposition versus community exposition?	3	V5
9. How many deaths resulted from COVID-19 in Canada?	2	
10. How do the various Montreal's regions compare to each other in the number of cases over time?	5	V3
11. How do the various Canadian provinces compare to each other in the number of cases over time?	3	V1
12. How do the various Quebec's regions compare to each other in the number of cases over time?	3	V2
13. What proportion of COVID-19 infected has recovered from each Canadian province?	3	
14. Which Canadian province is most affected by the COVID-19 pandemic?	2	V1

6. Mockups

This section presents different design mockups suggested to address the previous target questions identified in section 5. Each mockup is illustrated with a sketch and accompanied by a justification of the design choice and a description of the possible user interactions.

6.1 Visualization 1 (V1)

Visualization 1 addresses the following questions:

11. How do the various Canadian provinces compare to each other in the number of cases over time?
14. Which Canadian province is most affected by the COVID-19 pandemic?

Visualization 1 presented on the figure below is composed of both a series of 1D heatmaps as well as a side choropleth map.

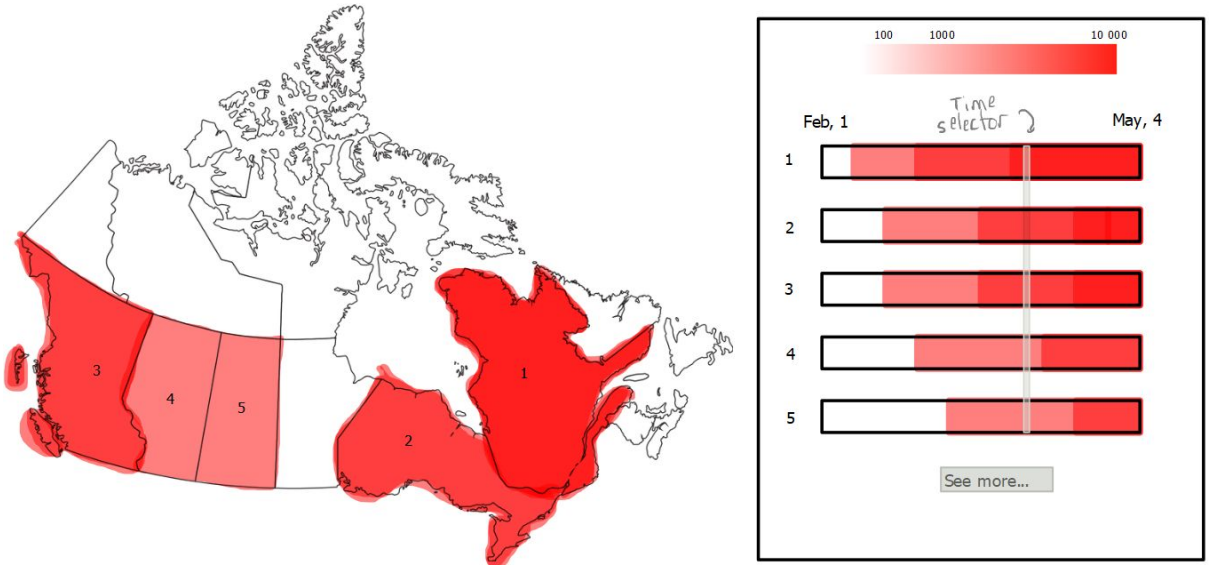

Figure 1. Mockup of visualization 1

The anticipated task of the user for this visualisation is to compare the evolution of COVID-19 throughout the different regions.The objective of the visualization is to assist the user with spatial reference with the map part and time reference with the heatmaps part. The different heatmaps on the right side of the visualization allow to efficiently display the number of cases of COVID-19 registered each day for the different Canadian provinces. The names of the provinces are to be displayed in the place of the numbers on the left of each heatmap channel as well as on the map. The linear progression of the data and color encoding of the quantity of cases makes for optimal time comparison inside one lane. Moreover, the alignment of different heatmaps make it fairly easy to compare densities of different Canadian provinces at the same point in time. The 1D heatmaps are displayed in a logical

order, highlighting the provinces racking up the highest case numbers at the top. The choropleth map positioned on the left side of the design provides relevant redundancy and helps to portray the differences observed at a chosen date from a truly geographical standpoint.

Upon displaying this visualization, the user is presented with a storytelling setup in which the time selector is scrolling from start to end date, resulting in a short animation of the evolution of the cases on the choropleth map. The step animation shows the different milestones of the propagation of the disease. When the step animation finishes,the user is then textually prompted to explore the visualization by dragging the time selector along the horizontal axis formed by the heatmaps in a timeline fashion and can compare cases between multiple regions on the map or along the verticality of the selector. Changing the position of the selector updates simultaneously the colors on the choropleth map for an eye-catching interaction. Moreover, the user can hover an area of the choropleth map to obtain the specific number of cases registered for that region, for a clutterless design. With the same intent, only a determinate number of heatmaps are shown directly. If the user wishes to, they can choose to display more information that will appear at the end of the heatmap list. The choropleth map is to move along according to the focused area of the user's scrolling so it is always displayed. The user has a filtering choice allowing them to choose a selection of regions to compare. This selection groups their chosen heatmaps at the top of the list and grays out the remaining heatmaps for close comparison.

The choropleth map may not be necessary in the case of the map of Canada because most of the target users are probably accustomed to identifying the different regions of Canada, but it becomes relevant for Montreal's or Quebec's numerous regions, that are addressed in following similar visualizations, that may not be known to all users. For repetition purposes and creating strong unity between V1, V2 and V3, the map was included in all of them.

6.2 Visualization 2 (V2)

Visualization 2 addresses the following question:

12. How do the various Quebec's regions compare to each other in the number of cases over time?

Visualization 2 presented on the figure below is composed of both a series of 1D heatmaps as well as a side choropleth map.

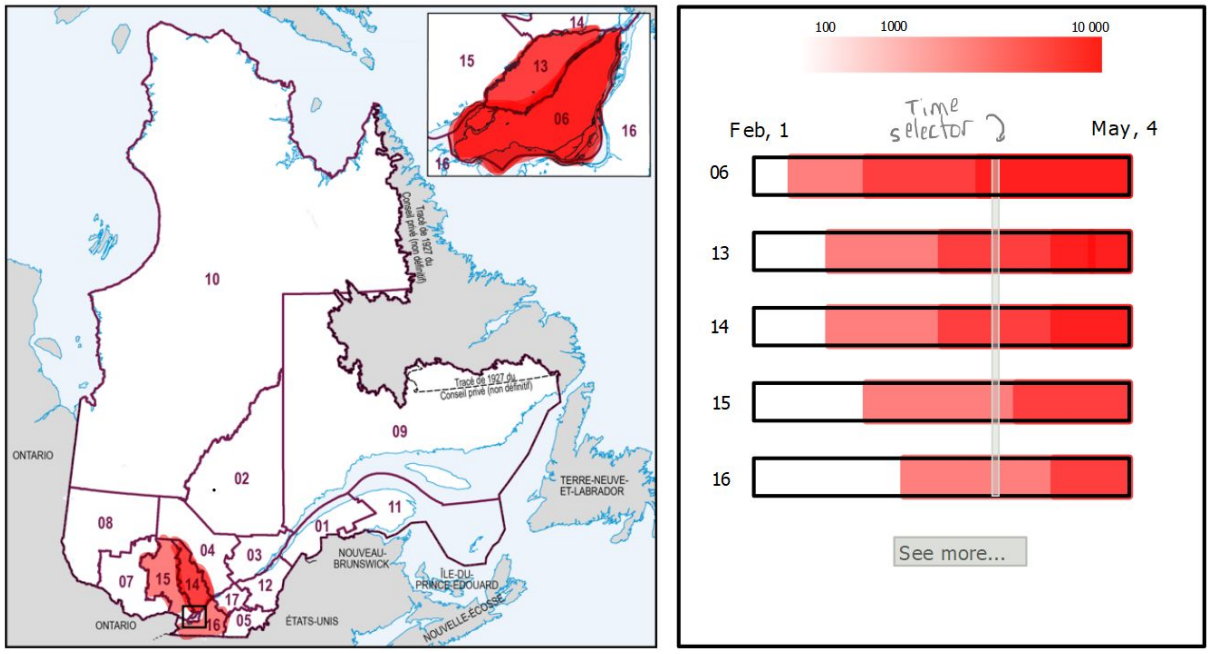

Figure 2. Mockup of visualization 2

Motivations for this visualization are essentially the same as presented for the visualization 1 in previous section 6.1. Again, this design corresponds appropriately to the target question.

6.3 Visualization 3 (V3)

Visualization 3 addresses the following question:

10. How do the various Montreal's regions compare to each other in the number of cases over time?

Visualization 3 presented on the figure below is composed of both a series of 1D heatmaps as well as a side choropleth map.

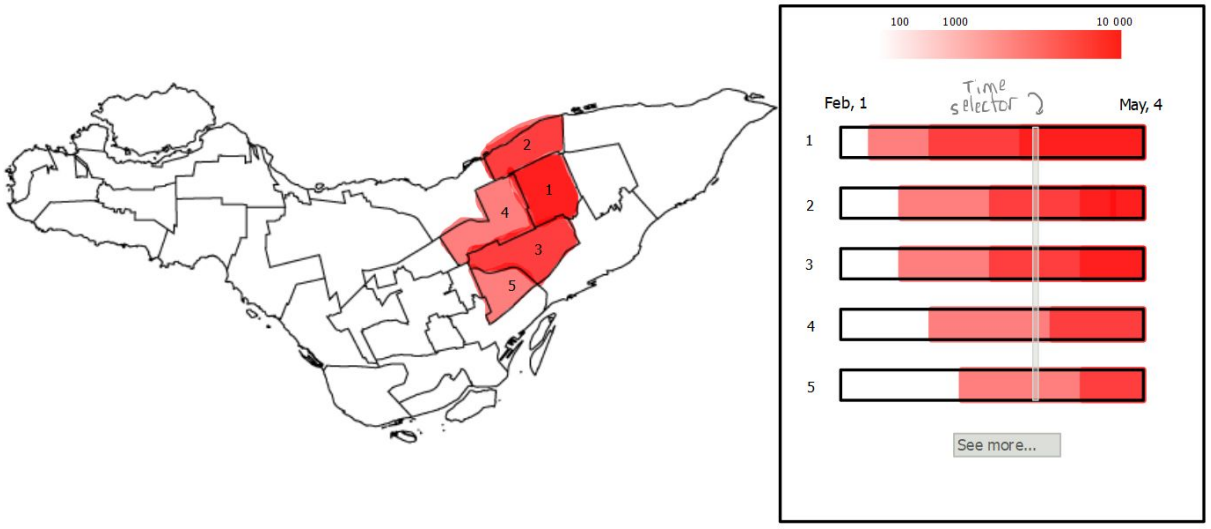

Figure 3. Mockup of visualization 3

Motivations for this visualization are essentially the same as presented for the visualization 1 in previous section 6.1. Again, this design corresponds appropriately to the target question.

6.4 Visualization 4 (V4)

Visualization 4 addresses the following questions:

1. Who was the age group most affected by COVID-19 in its early stage?
2. Who is currently the age group most affected by COVID-19?

Visualization 4 presented on the figure below is a dumbbell plot.

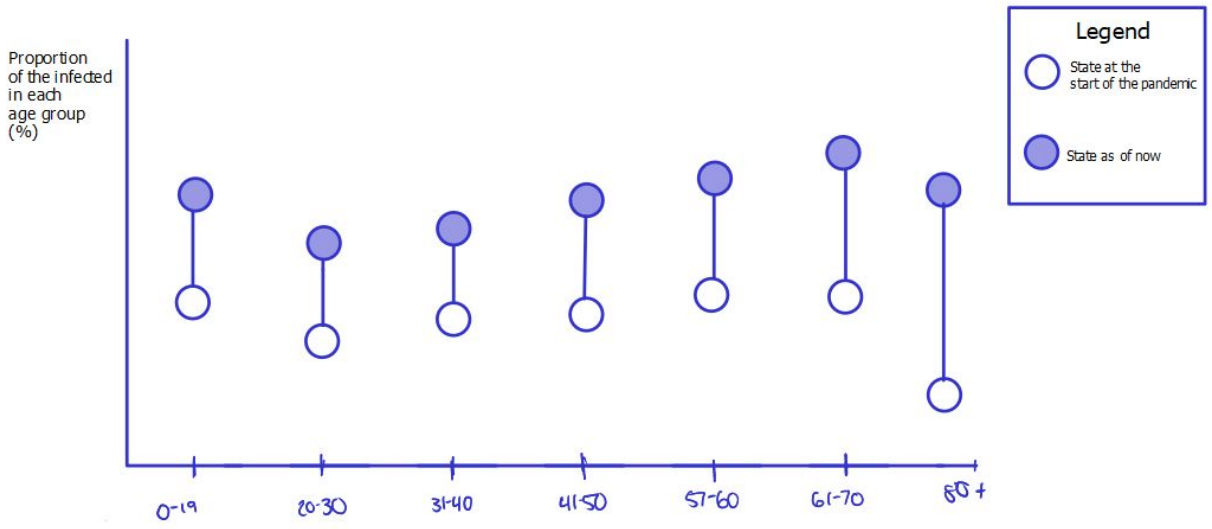

Figure 4. Mockup of visualization 4

Visualization 4 uses a dumbbell plot which is optimal for displaying differences between two main stages of the pandemic for different categorical dimensions, in this case the age groups, and shows comparison between the situation at the start of the pandemic and the current situation in a simple way. The dumbbell plot effectively allows for a very quick and precise view of the changes that occured in the displayed categories. The connecting line between the two categorical markers help the user understand which category has registered the biggest change over the selected period. The user is thereby able to see whether an age group is more affected at the beginning of the pandemic and whether it is still the case.

For this visualization, the user is presented once again with a scrollytelling kind of animation. The user is presented with the first set of dots representing the proportions of the infected that are from each group at the start of the pandemic. Then, the bars and the second set of dots are added upon scrolling to show the evolution of the proportions. This simple interaction prompts the user to compare the start of the pandemic to the most recent data.

6.5 Visualization 5 (V5)

Visualization 5 addresses the following questions:

3. Which gender has more complications in COVID-19 cases?
5. Are men more infected than women?
7. What proportion of cases required intensive care in Canada?
8. What proportion of the infections is transmitted by travel exposition versus community exposition?

Visualization 5 presented on the figure below is a Sankey diagram.

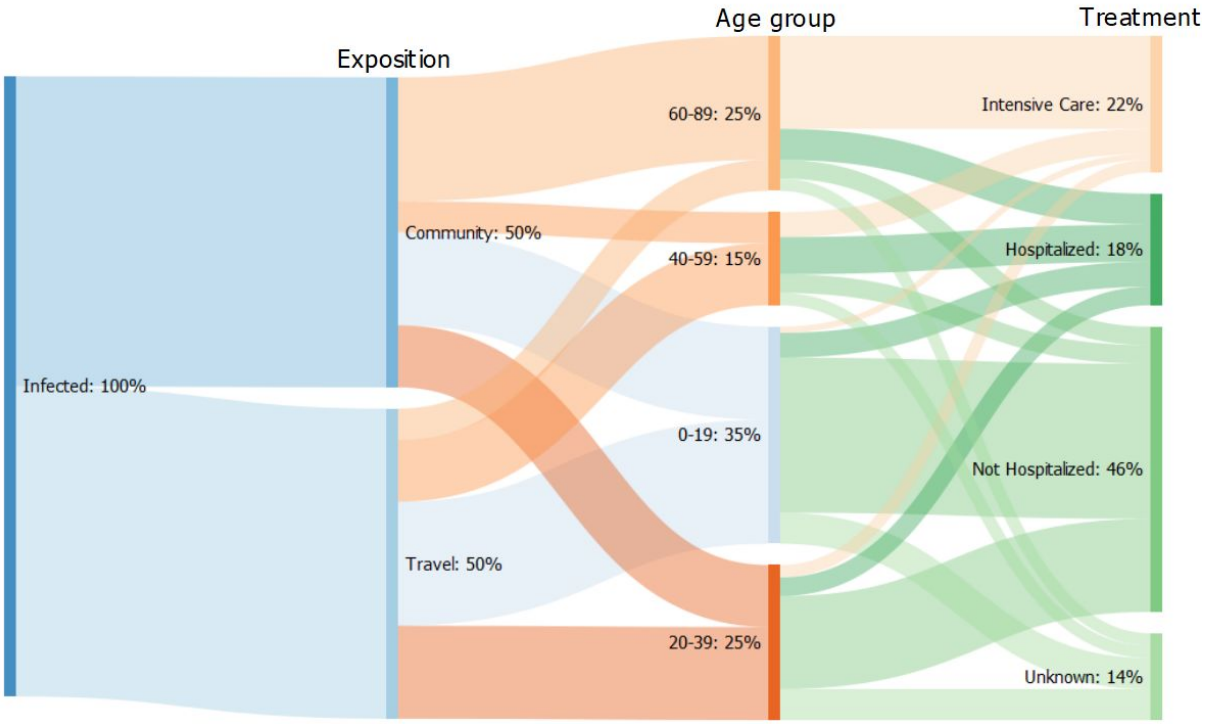

Figure 5. Mockup of visualization 5

Visualization 5 uses a Sankey diagram which is optimal for displaying categorical composition of the listed COVID-19 infected's profile which is questioned. The low number of categories at each stage allows to benefit fully from the Sankey visualization. Indeed, the different stacked bar charts are easy to read for categorical values comparison within the same stack, the bands' thickness is clear to the eye and the colors are distinct enough, as presented in the mockup. Quantitative flows of the curved bans link are also kept less intricate with the logical progression of the categorical dimensions in a whole to many guidance achieved by the chosen order: infected-exposition-age group-treatment. This order is also a meaningful choice since the relationships infected-to-exposition, exposition-to-age and age-to-treatment are probably the most interesting for users. Since the diagram is not

too crowded, labelling of the categories that make up the stack can be combined with percent information so as to provide more accurate quantitative data.

For this visualization, the user is presented again with a scrollytelling kind of animation. At first, only the stacked bar chart on the left accompanied by its label is presented. The second stack is then revealed with the user's scrolling and the flows from the left to the right sides are animated. The movement is ideal to entertain the user and highlight the flows' patterns as well as natural with the design choice. The following categorical dimensions are unveiled in the same fashion, going from general to more specific information on the COVID-19 infected.

7. Overall Design

The different visualizations proposed are to be organized in a scrollytelling overall layout paradigm. Since *Le Devoir* already has a dashboard and wishes to use as much data as possible to give a global comprehension of the evolution of the pandemic, this manner seems fitting. The anticipated layout will transition between the different angles proposed for the objectives from comparison between geographical regions to comparison between social groups to comparison between COVID-19 infected in a general to specific guidance. It could be possible that *Le Devoir* will wish to populate the space with descriptions and help to achieve these transitions with textual information highlighting important data displayed in the visualizations, which would mix well with the scrollytelling choice.

Concerning the organization of the visualizations 1, 2 and 3 which are all based on a similar heatmap and choropleth map mix, usage of the 3 visualizations could prove interesting with a tab layout since comparisons between the 3 different geographical levels seem unlikely. This would also prevent visualization redundancy for the eye and signal a similar visualization behaviour for easier exploration.Zooming animation going from one level to the other and updating the corresponding subregions heatmaps has been dismissed because of the cognitive charge it puts on the user that wants a fast switch between layers.