Communicating on Multivariate and Geospatial Data supported by ergonomics criteria: COVID-19 case

Hugues Turbé [1,2,*], Victor Garretas Ruiz [1,2,*], Mina Bjelogrlic [1,2,*], Jessica Rochat [1,2], Christian Lovis [1,2]

[1] Medical Information Sciences Division, Department of Radiology and Medical Informatics, University of Geneva, Switzerland

[2] Medical Information Sciences Division, Diagnostic Department, University Hospitals of Geneva, Switzerland

* these authors contributed equally to this work

Abstract

Since Fall 2019, the rapid spread of SARS-CoV-2 virus has changed everyday life routines globally. Public health confinement measures have been taken to contain the propagation of the pandemic. An international effort has been made to model and predict the spatio-temporal evolution of the pandemic . Today, a main question arises on how to communicate complex multivariate, geospatial and time dependent information efficiently. A further challenge consists in communication these information without any bias or place for misinterpretation, and for the largest targeted audience. In this regard the following paper will first identify ergonomics criteria for efficient data visualization, and then present several visualizations in a pre/post fashion, reflecting how visualizations initially proposed by data scientists can be improved after the application of ergonomics guidelines.

Keywords: Data visualization, Ergonomics, Infectious Diseases, Health Sciences, Epidemiology, SARS – CoV-2, Public Health.

Index Terms: K H.5.2: User Interfaces, J.3 [Life and medical sciences], E.2 [Data storage representations]

1 Introduction

Since its apparition in December 2019, the Covid-19 has had a significant impact on all aspects of society including a major strain on most health systems as well as major economic disruptions. The crisis was declared a pandemic by the world health organization[1] on the 11th of March 2020 and required governments to take Non-Pharmaceutical Interventions (NPIs) such as closing schools and banning public gatherings. With these measures, all citizens inherited some responsibilities for the future trajectory of the pandemic and became more interested in the evolution of the pandemic. Expressions such as “flatten the curve” and the reproduction number became part of the common language.

An extensive amount of data related to the pandemic has been published in open access with most countries publishing the daily number of confirmed cases, while others have also been publishing the number of tests carried out. This unprecedented amount of data lead to several visualizations including infections and death curves becoming common in the media. Visualizations have been critical for the public to make sense of the evolution of the pandemic and governments have extensively used a range of visualizations[2–4] to justify the different restrictive measures.

Throughout the pandemic, the following four dimensions were deemed critical to fully characterize the state of the pandemic in a given population: i) speed: number of new infections in a given time; ii) acceleration: change in the speed of infections over a given time; iii) uncertainties: relative to testing policies, reporting of the confirmed new cases, new deaths; iv) spatiality: information at different scales as well as the interactions between the different geographical entities.

Visualizations are powerful tools to organize information and show trends[5], but they can become complex when several dimensions need to be represented[6]. To minimize the complexity of the developed visualisations and favor their understanding, both by decisions makers as well as lay people, we used ergonomics criteria to evaluate the usability. Carpendale et al.[7] identified three generic requirements for visualizations which are generalizability, precision and realism. Further researches aimed to develop more specific criteria for the evaluation of information visualization techniques focusing both on usability and visual representation. Luzzardi et al.[8] developed their own criteria and compared them with the ones developed by Bastien et al.[9] and Nielsen[10]. Pillat et al.[11] showed that developing information visualizations with the ergonomic criteria developed by Bastien et al. facilitated the analysis and interpretation of the visualization. Huan et al.[12] took a different approach evaluating visualizations by measuring the cognitive load required to understand a given set of data. Concerning, ergonomics criteria that focus on visualizations, Kosslyn proposed specifics criteria based on human information processing [13]. Given their specificifty, those criteria were favored rather than more general criteria which apply to interfaces in general.

In this report, we aim to describe how a collaboration between data-scientist and user experience (UX) experts helped shape a set of visualizations relevant to describe and understand the evolution of the pandemic for scientists, governments and citizens. Firstly, this paper presents a set of indicators deemed critical to assess and compare the evolution of the pandemic across different geographical entities. Secondly, relevant ergonomics criteria used to evaluate the visualizations from an UX perspective, and the corresponding modifications made to meet this criterion are discussed.

2 METHODOLOGY

2.1 Building a set of relevant indicators

Given the unprecedented amount of data published in open access, our aim has been to evaluate the most relevant information, merge the different sources of information in an unified database and then bring those data to the broader public by producing compelling visualizations. The set of data used for the visualizations are presented in Figure 1 and the data are categorized following their level of abstraction.

Raw data consists of indicators which are easily understood by the vast majority of the public but have limited values when they are compared between geographical entities. Two levels of abstraction are built on these data to allow relevant comparisons. However, as these data become more abstract, they require a greater effort to be understood and this difficulty need to be factored in when building visualizations.

HIGHER ABSTRACTION →		
Raw data	**Calculated data**	**Estimated data**
-Number of tests -Demographic data -Confirmed cases -Confirmed deaths	-Incidence -Stringency -Positive rate	-Reproduction number -Infections
COUNTRY DEPENDENT ←		

Figure 1: Abstraction and dependencies of indicators

2.1.1 Raw data

Toward the start of the pandemic, confirmed cases obtained via PCR tests as well as deaths were aggregated by Johns Hopkins University[14] to meet the growing need for public data. The number of tests carried out in a given country has been aggregated by Our World in Data[15], but is available for a much smaller set of countries. The later source also provides demographic data which are required to calculate the incidence.

2.1.2 Calculated data

The incidence measures the number of new cases or death per inhabitants and allows a fairer comparison between geographical entities with different populations. To evaluate the impact of confinement measures on the evolution of the pandemic, the stringency index, developed by the Blavatnik School of Government as part of the Coronavirus government response tracker (OxCHRT)[16], was used. This index introduces a global score for each country based on the severity of NPIs taken including for example restrictions on gathering size or internal movement.

2.1.3 Estimated data

Estimated data include indicators dependent on properties which are intrinsic to the disease. These properties are evolving as the common level of knowledge regarding the virus grows. The date of infections is derived from the date at which confirmed cases or deaths are reported accounting for the delay between the infections and the reporting of these events. This step is crucial to consider the delay between the introduction of NPIs and a change in the level of infections or deaths. To infer the date of infections, distributions between the date of infections and the time at which a case is confirmed or the death of an individual are required[17]. The reproduction number measures the number of people infected by a single infected individual. A reproduction number above one indicates an acceleration of the pandemic in the observed population while the opposite is true for a reproduction number below one.

2.2 Expert evaluation of visualizations

An expert evaluation, also called "heuristic" evaluation refers to the usability evaluation of an interface by one or several experts in the field, based on ergonomics criteria[18]. This method is very beneficial when used in the early stages of design since it allows to reduce the number and severity of usability problems in an interface before user testing. An expert evaluation method, usually used to assess interface, was carried out to evaluate the graphical visualization made by data scientists. The evaluation was conducted using ergonomics criteria defined by Stephen M. Kosslyn that apply specifically for graphics and visual representations[13]. Kosslyn notes that if graphics are intended to be read by humans then this requires an understanding of the human visual information processing. Kosslyn has therefore classified these criteria following the three phases of the human visual information-processing system[13].

2.3 Ergonomics criteria

2.3.1 Getting Information into the System

The first phase involves how humans perceive what they see and includes four criteria: i) **Adequate Discriminability**: Marks must have a minimal size and must be discriminable; ii) **Visual Properties**: Some properties are better than others for conveying information (size, orientation, darkness, hue, intensity, texture etc.); iii) **Processing Priorities**: Humans first detect some differences in visual properties (e.g. brighter colors are detected before dimmer ones) considering them more important; iv) **Perceptual Distortion**: Distortion can lead to misjudgments.

2.3.2 Short-Term Memory Constraints

The second phase involves the limited short-term memory of humans and includes two criteria: i) **Perceptual Grouping**: Aggregating elements in an inappropriate manner can lead to misinterpretations as stated in the Gestalt laws of organization.; ii) **Memory-Capacity Limitations**: The short-term memory may be saturated if the graphic is overcrowded.

2.3.3 Long-Term Memory Processing

The third phase concerns the human's prior knowledge to recognize what he sees and includes two criteria : i) **Ambiguity in Labels and Design**: Ambiguity can occur if titles and labels are confusing and if it is not clear how the different parts of the graph fit together; ii) **Inferences**: Humans can make bad associations by looking at a graph leading to a misleading interpretation.

Three more general criteria are also described influencing the choice of representations based on: i) Purposes; ii) Types of question; iii) Data and formats.

2.4 Focus groups: data scientists and ergonomists

Focus group is a method which allows collecting participant's thinking on a specific topic[19]. Several rounds of focus groups were organized to improve the presentation of the different graphs produced by the data sicentists. Each round focused on one graphical representation of the data and included two ergonomists and two data scientists. During the focus groups, discussions were held to find solutions regarding the different problems detected using Kosslyn's criteria and to find the representations which were the best suited to convey the desired information.

3 RESULTS

Several visualizations depicting the Covid-19 evolution were produced across different rounds of focus group. The results are organized into three case studies with the final visualizations shown in this section, while the different versions produced during the generation process are shown in Appendix A-C.

Case study 1: Public health measures impact on SARS-CoV-2 evolution. The objective of Figure 2 is to show the relationship between the evolution of confirmed cases, R_t, and the stringency measures taken by the government.

Case study 2: Comparison of the incidence across countries. Figure 3 shows the ratio of positive tests as well as a measure of the uncertainty on the testing policy.

Case study 3: SARS-CoV-2 dynamics. Figure 4 shows the relationship between R_t and the mean infection per 100'000 people. The objective of this graph is to relate the evolution of R_t with the number of infections.

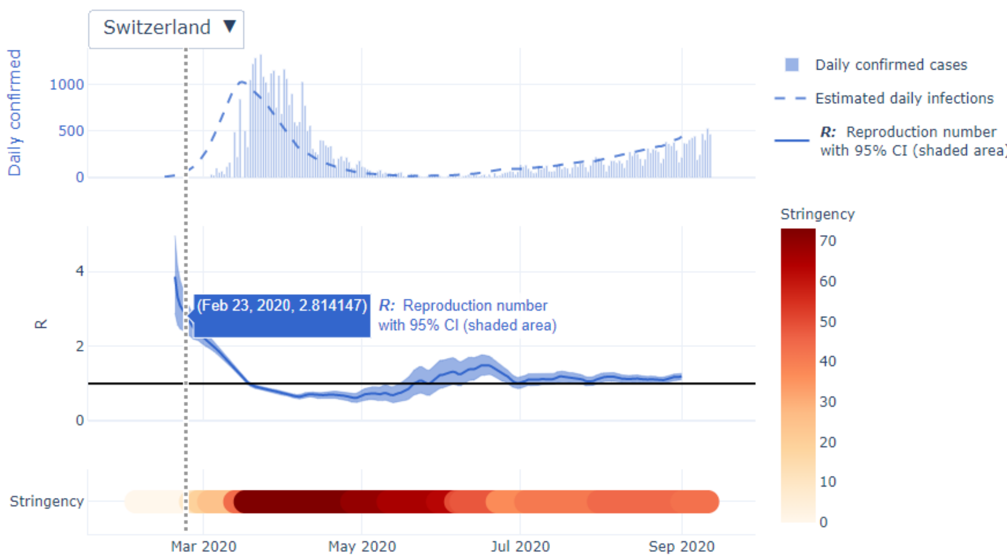

Figure 2: Relationship between the number of confirmed cases, R_t, and the stringency index (v2 after the initial focus group)

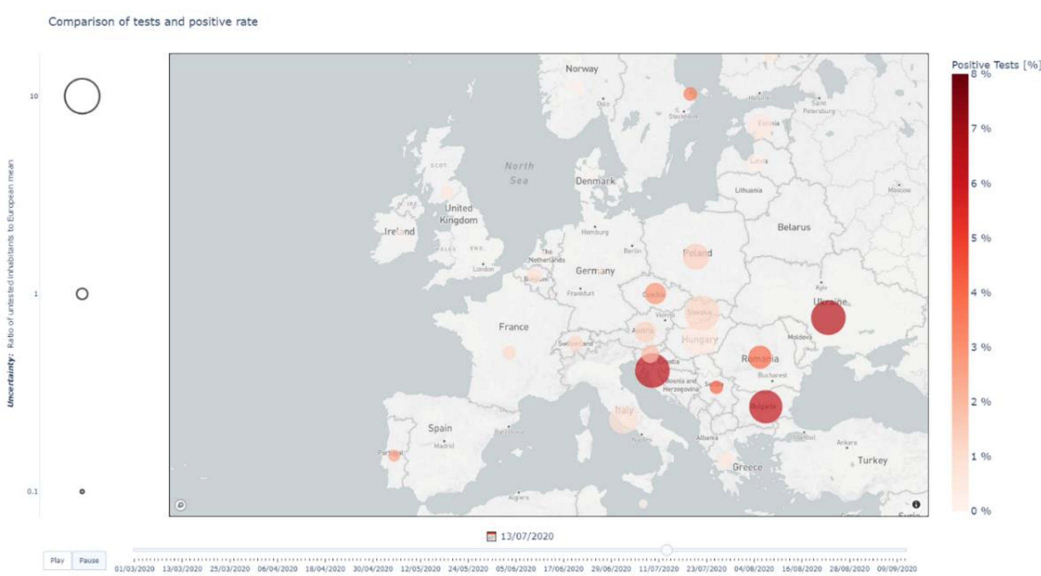

Figure 3: Comparison of the positive rate and the incertitude around those tests (v3 after the second focus group)

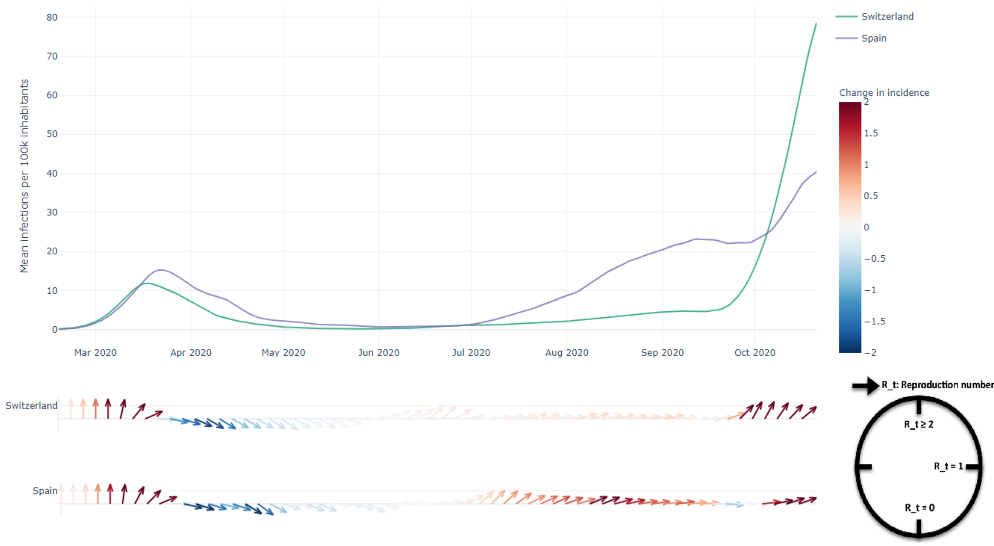

Figure 4: Impact of R_t (v3 after the second focus group)

4 Discussion

4.1 Evaluation of indicators pertinence

Towards the start of the pandemic, visualizations focused on the number of confirmed cases and deaths with the different infections' curves being compared across countries. However, the number of new infections as presented in Appendix A (Figure A1 and A2), with only the number of confirmed cases does not reflect the incertitude created by countries performing a different number of tests. The level of confirmed cases is highly dependent on the testing policy. Similar issues regarding the death count are encountered with some countries only reporting deaths in hospital and not in elderly care homes. Communicating on the incertitude linked to the different variables is critical to maintain the trust of the reader and therefore maintain public adherence to the imposed NPIs. This incertitude is represented in Figure 3 by showing not only the positive rate, but also the number of tests being carried over a given period.

The reproduction number also allows comparisons between geographical entities with different testing policies or reporting. While the R_t does not allow a comparison of the level of infected people at a given time between countries, it allows to compare the acceleration of the pandemic across countries. This factor was particularly relevant towards the start of the pandemic, when a range of NPIs were introduced and it was important to understand their impact on the evolution of the pandemic as shown in Figure 2. The reproduction number, R_t, however has several drawbacks. First it is dependent on changes of testing policy within a given country. Indeed, if a country increases the number of tests carried out, the number of positive cases reported will inevitably increase without reflecting an acceleration of the circulation of the virus. This factor can be mitigated by computing R_t on the number of deaths reported. However, this leads to another drawback, that is the reproduction number becoming inaccurate when calculated on exceedingly small number as shown in Appendix A, Figure A1 (middle graph).

Interpreting the reproduction number without the level of current infections can be misleading. Indeed, two countries may have a similar R_t, but vastly different level of infections putting them into two quite different situations as illustrated in Figure 4. While Switzerland had a slightly higher R_t than Spain in early September, it had a much smaller incidence and was therefore in a less critical situation.

4.2 Ergonomic evaluation

Solutions for including the inference, ambiguity, and memory-capacity limitations criteria are discussed in the following part.

4.2.1 Inference criterion

Case study 1. The two curves presented in the top graph of Figure A1 in Appendix A are difficult to discern. The blue curve represents the number of confirmed cases while the purple curve represents the number of deaths related to Covid-19 infections. The inference criterion is not respected and the visualisation might be interpreted as suggesting that there are as many deaths as confirmed cases. This happens because there are two different scales on the Y-axis which was identified as carrying some risk of misimpressions by J. Bertin[13]. To solve the problem related to inference problems, it was decided to remove the death variable and keep only the number of confirmed cases.

Case study 2. Although Figure A2 is more readable than Figure A1, it suffers from the inference criterion and conveys the wrong message. Looking at this graph one could wrongly deduce countries with the highest level of infection, because it creates an unfair bias towards countries with large testing policies. The "Processing Priorities" criterion was used to tell the reader which countries are in a critical situation. Indeed, the degree of uncertainty is represented by a circle on each country which increases to represent a larger uncertainty. The positive rate is represented by shades of red, with darker shades implying a higher positive rate. As this criterion says that the human sees first the big and brighter elements, the readers are attracted first to countries with high positive rate and low level of testing.

To represent countries that could potentially underestimate their level of infections over time we used an animation. The choice of an animation for this graph seems relevant to us since the purpose of this graph is not to see a trend but to see how each country evolves and compare them at a given date. It is possible to stop the animation and search for a particular date using the slider below the map to compare the countries.

4.2.2 Memory-capacity limitations criterion

Case study 3. While being difficult to understand at first glance, Figure C1 in Appendix C is powerful as it depicts both the current level of infection as well as the acceleration of the pandemic. Uncommon visualizations can work well if readers learn how to interpret them[20]. On this graph, each day is represented by a dot. The smaller dots represent the oldest days, with the size increasing for the most recent dates. 30 days per country are displayed. It is then possible to animate the graph moving the slider to show the following days. It is thus possible to see the evolution of the number of cases and the corresponding R_t. The speed at which the number of cases evolves is represented by the spacing between the dots. The more the dots are spaced, the greater the speed. Acceleration is represented by the fluctuation of R_t . The drawback of this graph, as for animations in general, is that it requires more cognitive effort as the old points disappears when the slider is used. It is thus difficult to see a trend since you must remember points that you no longer see. This is the reason why it is generally recommended to use timelines rather than animations. However, the disadvantage is that timelines, as shown in Figure C2, scale badly on larger time periods as they become overcrowed[21].

4.2.3 Ambiguity criterion

Case study 1. In Figure A1, showing both the number of confirmed cases and the number of deaths creates an ambiguity in the top graph. The bar charts represent the actual measurements while the curves represent the estimated date of infection. Mixing the bar chart representing both the number of case and death does not respect the ambiguity criterion as both bar charts are juxtaposed. As such the visualization might suggest that we are measuring proportions with the juxtaposition of bar charts looking like divided bars. This juxtaposition problem is also present in the middle graph.

Moreover, it is hard to understand that the bar charts represent actual measurements while the curves represent the estimated date of infection. This problem occurs because the ambiguity criterion is not respected. It is ambiguous because the items in the legend are unclear or missing not allowing the reader to understand. In addition, it is also not immediately understood that the three graphs share a common X-axis which lies in the bottom graph. In Figure 2, removing the death variable solves the problems related to the ambiguity criterion as there is no longer any juxtaposition of bar charts in the top chart nor any juxtaposition between the areas of uncertainty of R_t in the middle graph. To solve the problem of understanding the information displayed in the graph, we modified and completed the legend. Indeed, it was not possible to guess that the bar chart represents the confirmed daily cases and that the curve represents the estimated number of infections without explaining it in the legend. To make it clear that the three graphs run at the same a vertical line across the three graphs which appears when we place the cursor over one of the three graphs was added.

Case study 2. In Figure B1 in Appendix B, it is not easy to compare the different countries as different colors are used to represent different quantities. It is not easy to answer the following questions: "Does France have more cases than Spain? Does France have more cases than Turkey?" without looking at the legend every time you want to compare two countries. This happens because the visual properties criterion is not respected. As said by Kosslyn: "Differences in quantities should not be represented by differences in color" because "shifting from red to green does not result in "more of something" in the same way as shifting from a small dot to a large one does". In this case study, it is possible to use Tufte's advice. "He suggests the use of varying shades of gray to represent ordered quantities"[13]**.** In Figure B2, the incidence is shown by varying the shade of a single color.

Case study 3. Figure C2 in Appendix C uses a timeline; however, the visualization is not easy because we need to represent more than two variables with only two axes. 3D visualizations have been excluded as many researchers have found that they decrease the understanding of the readers[20]. The third variable was represented by a "retinal" value as said by Bertin in Kosslyn's review[13]. The time dimension is represented in the X-axis, the R_t in the Y-axis, and the mean infections is represented by the size of the dot. This graph shows well the evolution of R_t over time, but it is difficult to see the mean infections as the size of the dots does not vary much. In addition, when the mean infection is low, the dots become exceedingly small, which may not respect the discriminability criterion. To improve the understanding of the relationship between R_t and the increase in the number of infections and to overcome the limitations of Figure C2 , a third graph shown in Figure 4 has been produced. We started from a classic line chart where the time is in the X axis and the average infection in the Y axis. In this way we avoid the discriminability criterion, and we reduce the ambiguity criterion. An additional plot is used to show the evolution of the R_t using arrows. Starting from the bottom, the arrow progressively points upwards as the R_t increases. This arrow shows the acceleration. The advantage of this visualization is that one can understand that when the R_t increases then the mean infection increases without understanding requiring an definition of the R_t . In order to show that acceleration does not have the same impact according to the level of mean infection, we have combined the arrow with color. The color shows the increase in mean infections based on the difference in the number of cases between four days. Thus, the darker the arrow, the greater the speed at which the number of cases increases. By combining the arrow and the color we can represent both the change in incidence and the evolution of R_t. to emphasize their relation.

5 CONCLUSION

Data visualizations are a powerful support to carry a story and make sense of the abundance of data. It is however important to note that either purposefully, for ideological reason, or unintentionally (for example through a lack of understanding of the data) they can be misleading. It is particularly important for this reason for the designers of these visualizations to think about the aim behind the latter as well as any implicit suppositions which can influence the realizations of these visualizations. Given the implications on the general population, it was deemed essential to use ergonomics criteria for the range of visualizations produced during the crisis in order to try to improve as much as possible their comprehension by the recipients of information. That is the produced visualizations should be as relevant for general leaders who must decide on the introduction/removal of confinement measures than for the general public without whom the decided containment measures cannot be effectively introduced.

First, a set of indicators were identified to characterize the epidemic. This set includes the reproduction number, the number of test and positive rate. Different visualizations based on these indicators, were then produced by experts in data science. The initial designs were then followed by several focus groups with experts in ergonomics to identify potential problems with each graph. This heuristic evaluation guaranteed that some weaknesses in the initial visualizations could be detected and resolved. To consolidate the validity of these graphs and their understanding, it would be interesting to complete this heuristic evaluation with the help of user testing.

Data Availability

All visualizations presented in this report are available from the corresponding authors on request.

References

1. Coronavirus (COVID-19) events as they happen. Accessed September 15, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen
2. Griffin AL. Trustworthy maps. *JOSIS*. 2020;(20):5-19. doi:10.5311/JOSIS.2020.20.654
3. Sopan A, Noh AS-I, Karol S, Rosenfeld P, Lee G, Shneiderman B. Community Health Map: A geospatial and multivariate data visualization tool for public health datasets. *Government Information Quarterly*. 2012;29(2):223-234. doi:10.1016/j.giq.2011.10.002
4. Kinkeldey C, MacEachren AM, Schiewe J. How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. *The Cartographic Journal*. 2014;51(4):372-386. doi:10.1179/1743277414Y.0000000099
5. Fekete J-D, van Wijk JJ, Stasko JT, North C. The Value of Information Visualization. In: Kerren A, Stasko JT, Fekete J-D, North C, eds. *Information Visualization*. Vol 4950. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2008:1-18. doi:10.1007/978-3-540-70956-5_1
6. Cleveland WS, McGill R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*. 1984;79(387):531-554. doi:10.1080/01621459.1984.10478080
7. Carpendale S. Evaluating Information Visualizations. In: Kerren A, Stasko JT, Fekete J-D, North C, eds. *Information Visualization*. Vol 4950. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2008:19-45. doi:10.1007/978-3-540-70956-5_2
8. Luzzardi P, Freitas C, Cava R, Duarte G, Vasconcelos M. An extended set of ergonomic criteria for information visualization techniques. In: *Proceedings of the Seventh IASTED International Conference on Computer Graphics and Imaging (Cgim-2004), Kauai.* ; 2004:236–241.
9. Bastien JMC, Scapin DL. Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*. 1995;7(2):105-121. doi:10.1080/10447319509526114
10. Nielsen J. 10 usability heuristics for user interface design. *Nielsen Norman Group*. 1995;1(1).
11. Pillat RM, Valiati ERA, Freitas CMDS. Experimental study on evaluation of multidimensional information visualization techniques. In: *Proceedings of the 2005 Latin American Conference on Human-Computer Interaction - CLIHC '05*. ACM Press; 2005:20-30. doi:10.1145/1111360.1111363
12. Huang W, Eades P, Hong S-H. Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective. *Information Visualization*. 2009;8(3):139-152. doi:10.1057/ivs.2009.10
13. Kosslyn SM. Graphics and Human Information Processing: A Review of Five Books. Bertin J, Berg WJ, Chambers JM, et al., eds. *Journal of the American Statistical Association*. 1985;80(391):499-512. doi:10.2307/2288463
14. Johns Hopkins University (JHU CSSE). COVID-19 Data Repository. Accessed May 23, 2020. https://github.com/CSSEGISandData/COVID-19
15. Beltekian D, Gavrilov D, Giattino C, Hasell J, Macdonal B, Mathieu E. Data on COVID-19 (coronavirus) by Our World in Data. doi:26/05/2020
16. Petherick A, Kira B, Angrist N, Hale T, Philips T. Variation in government responses to COVID-19. *Blavatnik School of Government Working Paper*. 2020;31.
17. Turbe H, Bjelogrlic M, Robert A, Gaudet-Blavignac C, Lovis C, Goldman J-P. Adaptive time-dependent priors and Bayesian inference to evaluate SARS-CoV-2 public health measures validated on 31 countries. *medRxiv*. Published online January 1, 2020:2020.06.10.20126870. doi:10/gg5wh5
18. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. Association for Computing Machinery; 1990:249–256. doi:10.1145/97243.97281
19. Krueger RA. *Focus Groups: A Practical Guide for Applied Research*. SAGE Publications; 2014.
20. Kosslyn SM. *Graph Design for the Eye and Mind*. Oxford University Press; 2006. doi:10.1093/acprof:oso/9780195311846.001.0001
21. Beck F, Burch M, Diehl S, Weiskopf D. The State of the Art in Visualizing Dynamic Graphs. *EuroVis - STARs*. Published online 2014:21 pages. doi:10.2312/EUROVISSTAR.20141174

APPENDIX

Appendix A: Evolution of confirmed and death, Rt and stringency index

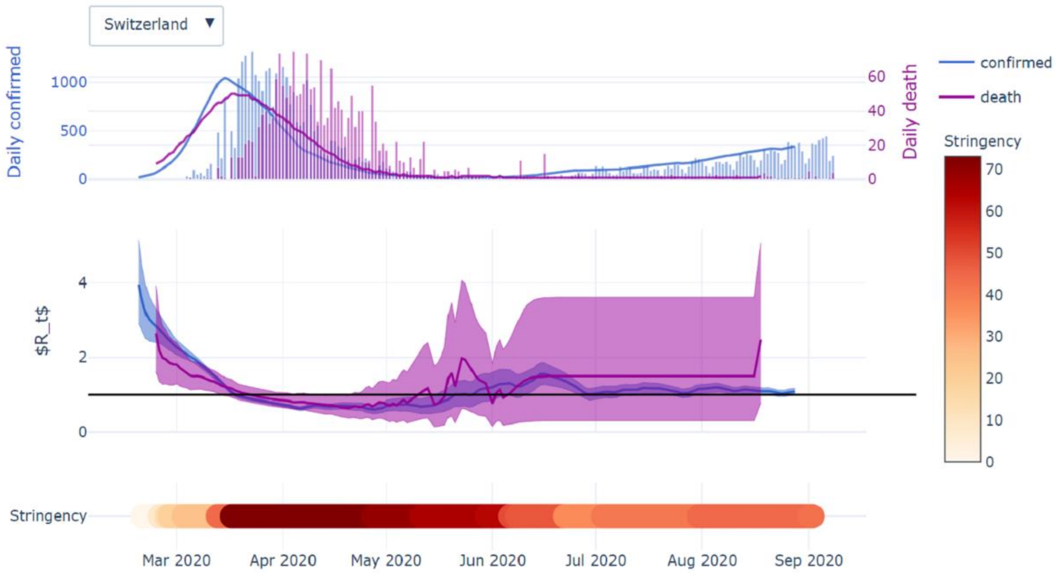

Fig A1: Relationship between the number of cases/deaths, Rt and stringency (v1: Before the initial focus group)

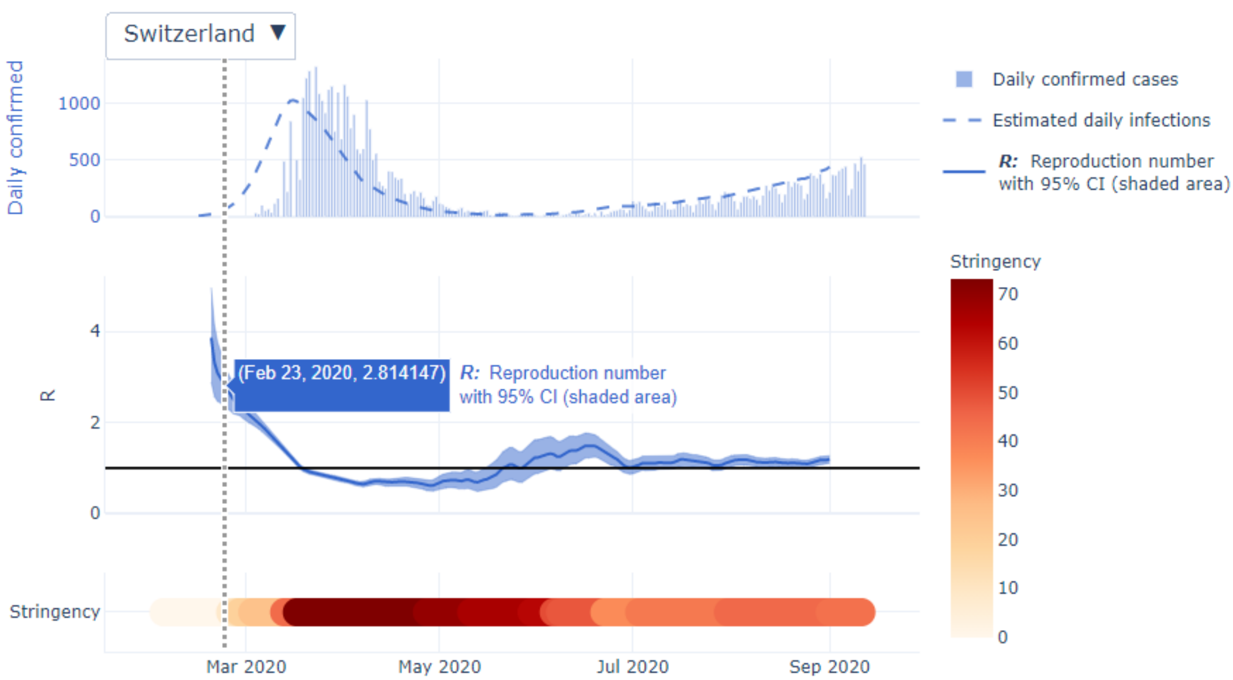

Fig A2: Relationship between the number of cases, Rt and stringency (v2: After the initial focus group)

Appendix B: Comparison of state of the pandemic across countries

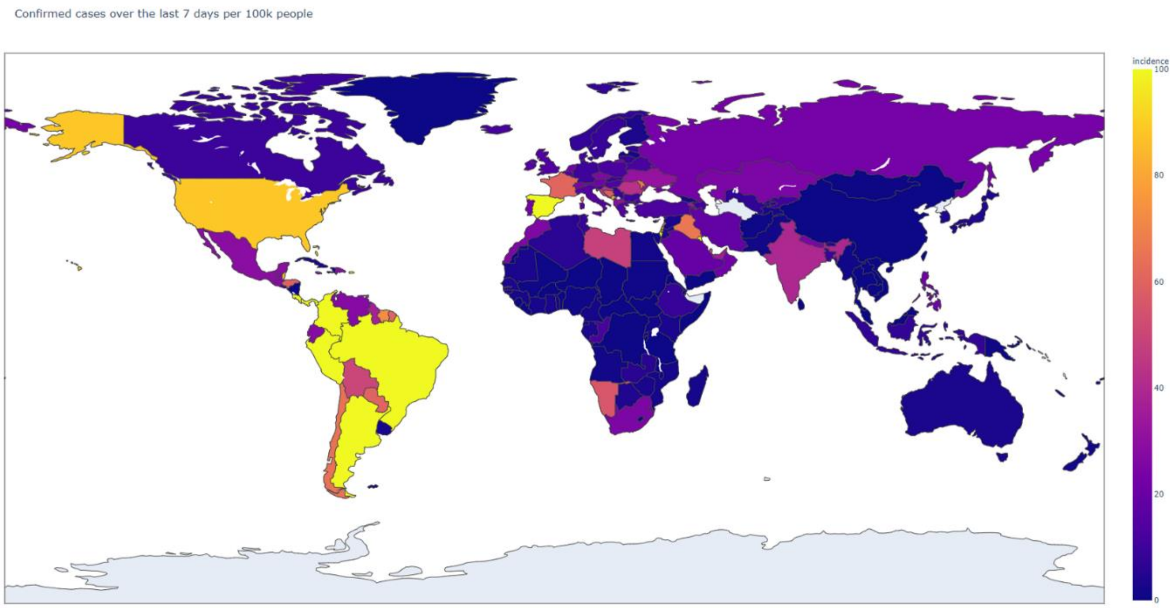

Fig B1: World map of the incidence (v1: Before the initial focus group)

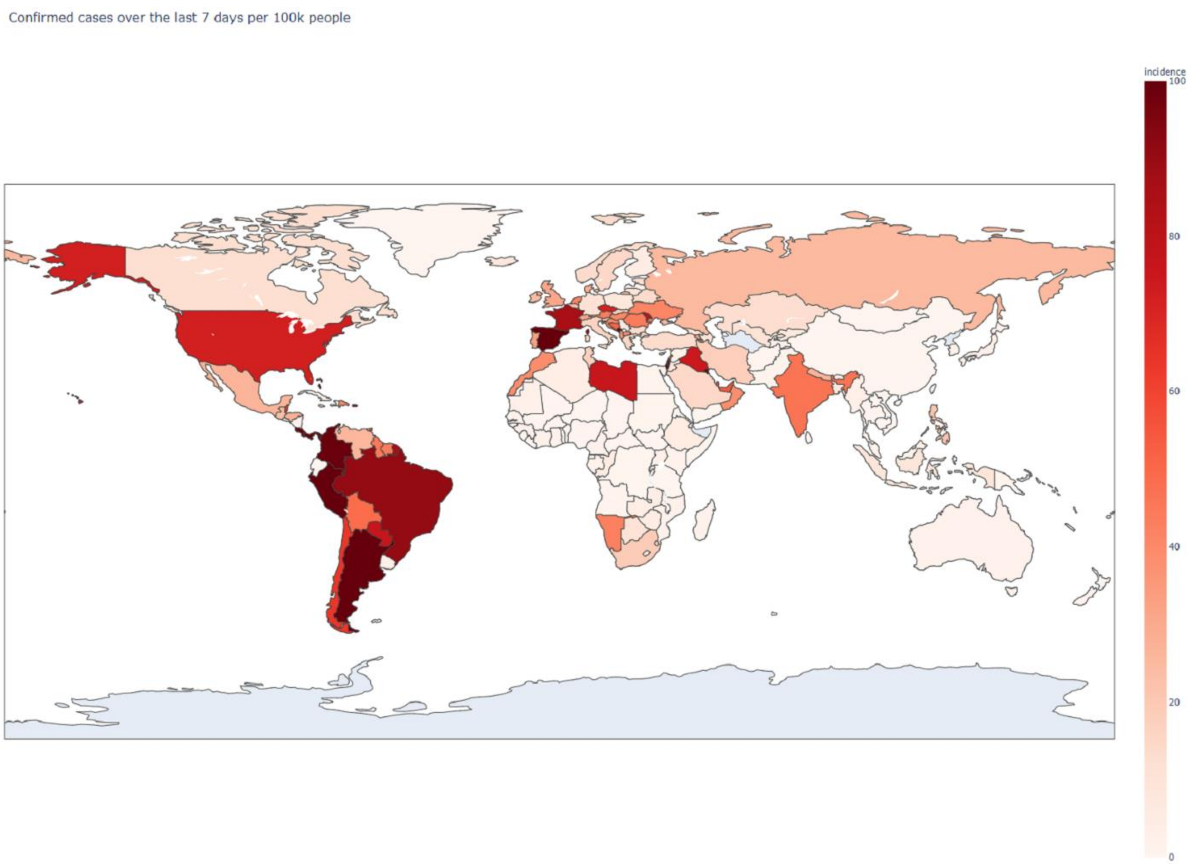

Fig B2: World map of the incidence (v2: After the initial focus group)

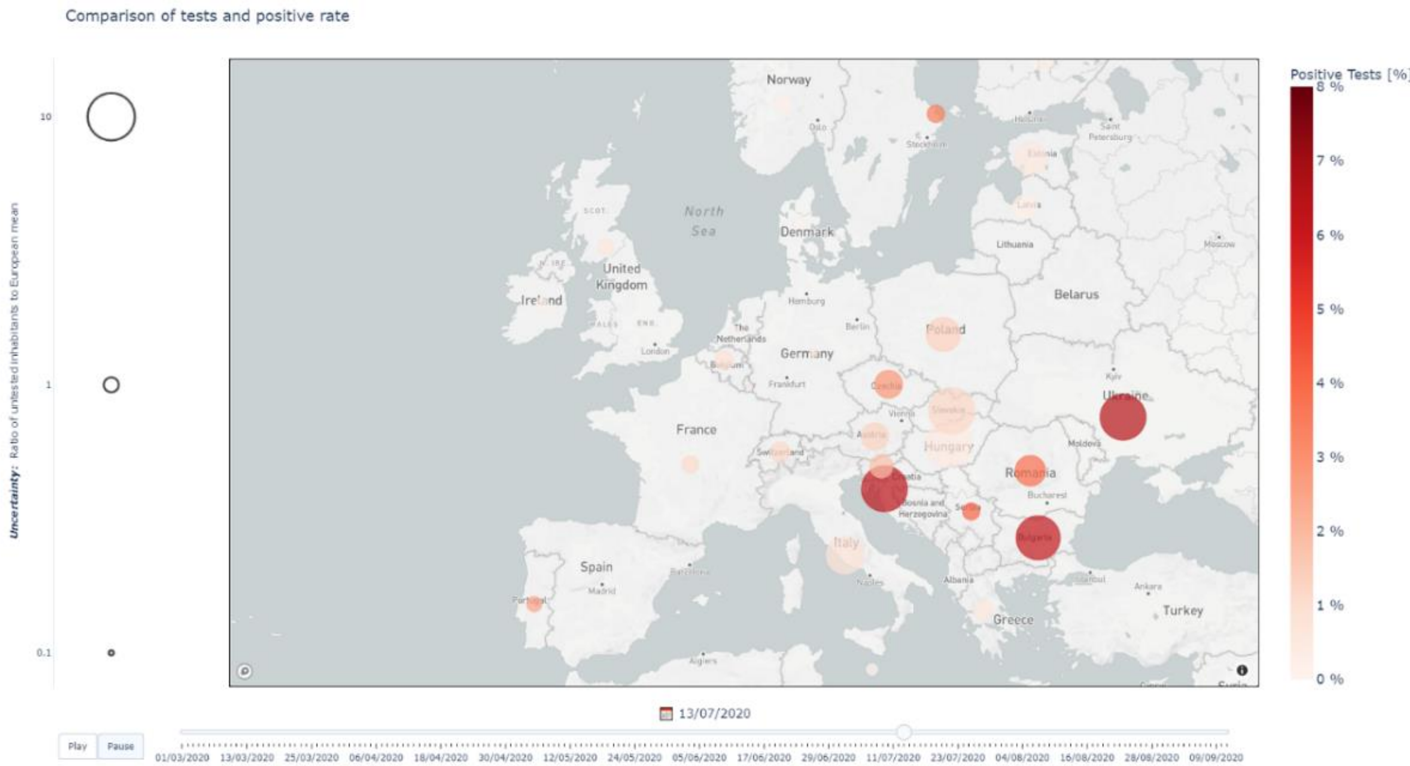

Fig B3: Comparison of the positive rate and the corresponding incertitude (v3: After second focus group)

Appendix C: Evolution of the reproductive rate and number of infections

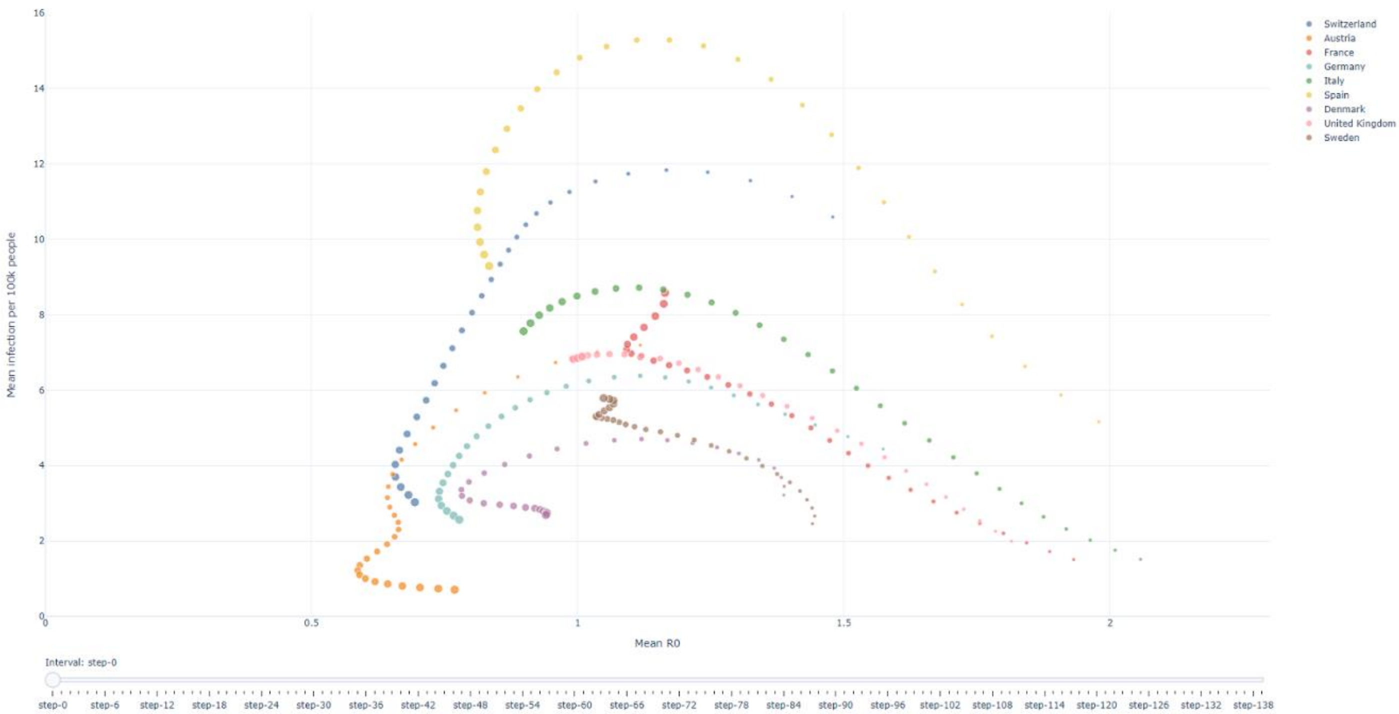

Figure C1: Mean infection vs. mean Re (v1: Before the initial focus group)

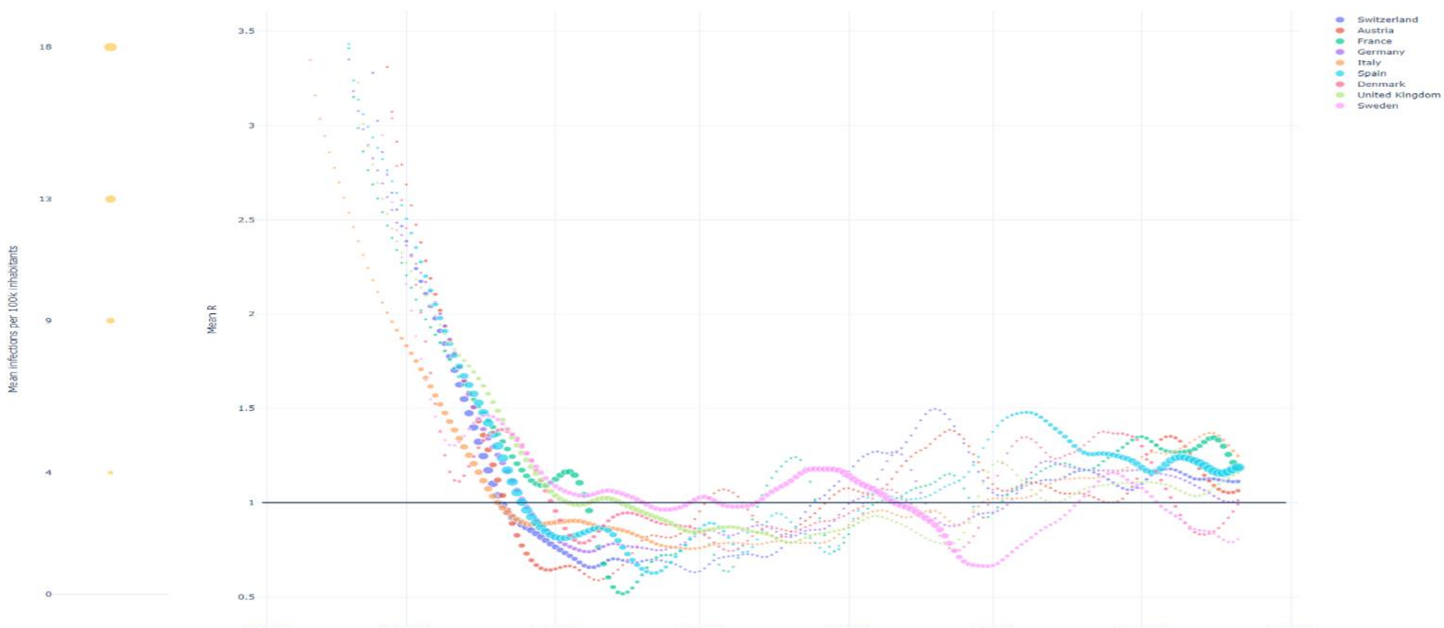

Figure C2: Mean infection vs. mean Rt (v2: After the initial focus group)

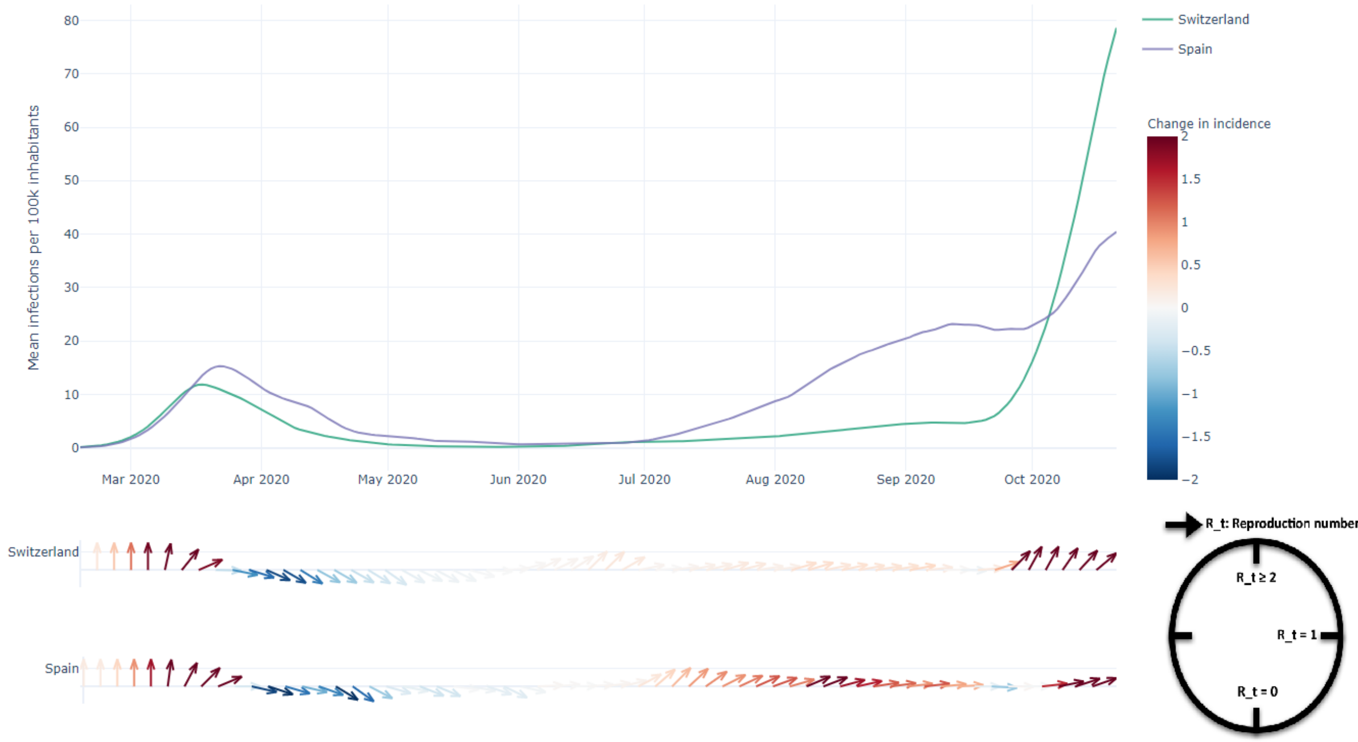

Figure C3: Evolution of mean incidence, Rt, and change in incidence (v3: After the second focus group)