Abstract

In this document, we are going to work on the parts three and four of the dashboard delivery. We will divide this document in two parts, the first one, regarding the selection of the charts and the encoding and the second one, regarding the implementation of the final Dashboard with Tableau. Before starting with the contents of this delivery, let us discuss the comments given in the previous delivery in order to improve our work.

1 Previous comments

Let us start re-describing the audience of our visualization. The first issue to take into account is that the audience are not (in general) experts in the use of charts. Actually, the audience can be any person interested in the most common topic of these two past years. Since this topic affects the entire population, anyone who wants to take a look on the topic can go through our visualization, no matter how old are they or how experts they are in the topic. Therefore, we can not make a detailed description of the audience. Regarding the question of how often will the audience consult the visualization, our idea is that, as long as the datasets are updated by public organizations, we will update the visualization, so they can consult the charts every time they need it. To simplify, the platform that the audience will use to see our charts is the computer, otherwise, we will have to rethink the sizes of the charts, position, etc.

We want to re-formulate the questions that the audience may want to answer while looking our visualization. The questions that we propose are the following: 'How the illness affected the health of the population by sex and age?, 'At which level the region or the density of population affected the spread of the epidemic?, 'How the economy influences in epidemiological issues? 'Is it more common than before, work from home? 'How much? 'How the illness affected the mobility of the population for every CCAA? More precisely, with this last question we would like to expose how the illness affected on the people's frequency to visit places like grocery markets, food warehouses and pharmacies compared to baseline. Our goal is to, with only looking the charts, take appropriate conclusions and answer the mentioned questions before and not only answer those, but present the information in a clear and visual way so that everything is fitted in a four chart dashboard.

Regarding the gender and age issue, we have decided to add another chart with this separation. Actually, we have rethought our dashboard, and we have changed some aspects. On the first screen, we display the most important information of the topic, and we answer the most relevant questions with mainly four charts concerning epidemiological issues. In this screen we will have a basic choropleth chart with new cases ratio per region information, two population pyramid charts with detailed information on new COVID-19 cases and new hospitalizations grouped by gender and age, and finally a global overview of new COVID-19 cases and new hospitalizations over the time in order to see the time evolution in one single chart, without the separation on gender and age.

The idea of this first dashboard is that when you click on a CCAA at the choropleth chart, all the information displayed by all the charts present on this dashboard will be filtered on the selected CCAA. Moreover, there will be a slide on quarters of time, working for all charts except for the one showing an overview of COVID-19 new cases and hospitalizations, since it is already represented over time. Therefore, the idea is to make a clear separation to the charts in which we can move over the time and to the global one, in which we cannot. Notice that with this first screen, the audience will be able to answer questions regarding epidemiological issues, such as how did the new cases of this disease evolve over time, how where they correlated to the different location frames, how was the hospitalization distribution over time during this period of pandemics, and so other natural questions that can be directly answered with the dashboard

we present.

In the second screen, we will find two basic choropleth charts and two combined area and line charts. In the first basic choropleth chart we will find information about the mobility for every CCAA and on the second one about the "work from home" for every CCAA, since it represents the mobility to the workplace. What we have mentioned will be on the upper part of the second screen, on the bottom part, we will have information about the unemployment rate and GDP (in area charts) beyond a line chart of new COVID-19 cases, following the same procedure as the aforementioned, since we also seek to compare the tendencies of both distributions qualitatively. As before, in this screen, when the audience selects one CCAA, all the information displayed will be regarding the selected CCAA. Moreover, there will be a slide on time working for the upper part of the screen, since again the bottom charts are already representing the values in terms of the time, so we will make a clear separation between those parts. Notice that with this screen, the audience will be able to answer the mobility questions and the economics questions we presented before.

To finish this part, let us mention the origin of the datasets used. For the economics charts, the ones containing information about GDP and unemployment rate, we have used two datasets provided by the INE. For the population pyramid chart regarding new COVID-19 cases by gender and age, we have used a data set comes from "Red Nacional de Vigilancia Epidemiológica"[1].For the rest of the carts, we have used the provided dataset of the subject, specifically, the epidemiology and the mobility csv.

2 Selection of chart and encoding

In this section, we are going to explain the type of charts and the encoding of the data, which was chosen taking into account the main goals and questions that our visualization should answer. Moreover, we will describe why we selected those charts and encode following the topics explained in class, such as perception, properties, principle and best practices. As already mentioned in the previous delivery, our dashboard consists of two screens of four charts each, where the main charts are those in the first screen. For sure there is plenty of COVID-19 information, and we could put more charts and try to extend them by adding further information to it to get more particular conclusions, but we stuck with the idea of minimizing the cost of visual searches, make the visualization displays as compact as possible, and we decided that four charts per screen are more than enough for today's instant click society. We also focused on presenting information about the pandemics on Spain, so the visualization will be closer to the audience, and would feel more represented by the data shown, since they have actually lived this situation and can relate most of the information to personal experiences and environment. This is also the reason of deciding to look into further data bases, so we would make sure we had enough information regarding the geographical location chosen.
The charts of the first screen are focused on health topics and how the epidemics spread along the population, whilst the second screen charts give detail of the socioeconomic situation.

We start by describing the first screen, where the epidemic information is given:

- On the upper left part, we can find a **basic choropleth** chart. It describes the ratio[2] of new cases in Spain per CCAA at different times that range from the first quarter of 2020 till the last of 2021. This chart will allow the audience to perceive at which level each of the regions is more wildly affected by the spread of the epidemic. We have chosen this as the main chart since it is the standard approach for putting data on a map, as well as, it is a really engaging method to start the visualization with.

[1] https://rubenfcasal.github.io/COVID-19/
[2] In order to avoid misleading information on the communities that had more population.

We found it interesting to see how our neighboring communities got infected by the virus as time goes on, we thought this would be to powerful and engaging for our audience too. In addition, we would like to remark, that not only we chose this kind of chart as the first chart because we found it important and interesting enough, but also because it is the one the user can interact with. Every other chart will be influenced by this, since as you click in a CCAA the other charts will be updated to see how was that region in particular dealing with COVID-19. A final commentary about this chart worth to speak of is the color intensity, we use a more intense blue as ratio of new cases increases.

- On the upper right part, we can find a **population pyramid** chart, in which we split the incidence of the epidemic, that is, the new cases at each step of time, by gender and by age. This chart, next to the basic choropleth chart, will allow the user to deepen the information given by the previous one. The idea is the following, as one is scrolling through time in the choropleth chart, the population pyramid also updates with that time to visualizes at which point was the population. Once the user is interested in a particular region and clicks on it, the pyramid gets updated and one can see a picture of the state of the population in regard to the incidence of the virus at that region and time. The goal we reach here is to show how the epidemic affected the health of each region by gender and age. We have chosen this kind of chart as it is the optimal way to visualize how the same topic affects different classes of the population. A little view to different regions and times in this chart, allow the user to get an idea of which group of age and gender tends to be the most affected by the virus. Thus, with the upper part of the dashboard, the audience can have, if wished, either a global vision, or a grouped one, filtering either by time, geographical region, gender or even age of the population, achieving a first objective of completeness, efficiency and transversal but at the same time specific vision of the information available.

- On the bottom left part, there is a different kind of chart, this is a combination of a **line** chart with an **bar** chart. It is different in the sense that we have the corresponding distributions of new cases and hospitalizations already in terms of the time, and thus, it is not affected by the time slider filter. However, it is still grouped by region, so it will update when the CCAA is selected on the choropleth. To separate it from the rest, it is highlighted and surrounded by a soft blue mark. The line chart represents the evolution of new cases throughout all time according to the selected region. We chose this kind of chart since it is the usual way to visualize time series. In the bar chart, the user can see the new hospitalizations along the time steps and also by selected region. We plotted both of them together, since our goal is to find correlations between these variables. One initial idea was also to include in this chart another line for the vaccinations to the three-way correlations, but we ended up discarding that option since we found out that vaccinations were not properly written in the dataset since they exceeded by many times the total population of Spain! The user can see, however, that hospitalizations increased at the very beggining of the epidemic and follow a similar pattern than new cases, until the second quarter of 2021. From that point onwards, hospitalizations are stabilized, even though new cases increase. It is particularly remarkable that this chart's objective is to see how the evolution of cases impacted the hospitalizations, globally and per region. Thus, we seek to compare the monotony and tendencies of both distributions. In order to do so, we wished to make a direct and clear chart that would transmit to the audience a difference between the cause, which is the confirmed cases distributions, represented by a line, and the consequence, which is the hospitalization distribution and is represented on the same chart as an bar chart. Note therefore that the domain ranges of both magnitudes are not synchronized, since this intends to be a qualitative chart of each distributions behaviour in comparison to the other, and would not be distinguished properly when using synchronized y-axis since the order of magnitude were significantly different.

- Finally, on the bottom right part, underneath the new cases' population pyramid, we can find another **population pyramid** but this time it displays the new hospitalizations. The idea is the same as the chart above. The goal is that the user can identify at which part of the population the epidemic was more intense, in the sense that who required medical assistance to deal with the symptoms. It is interesting to see the contrast of the two pyramid populations. We decided to place one over the other so that the user can compare them easily and distinguish the correlations faster.

Notice that, in the first screen, the audience will be able to answer the most important questions, with the basic choropleth chart the audience will be able to perceive at which level the region or the density of population affected the spread of the epidemic. Finally, with the population pyramid charts and the combined line and are chart, the audience will be able to perceive how the illness affected the health of the population by sex and age.

Now we are going to focus the explanation on the charts of the second screen, where the user can visualize the influence of COVID-19 on the socioeconomic behavior. We can easily distinguish two parts. On the top, one can find the correlation of the epidemic and mobility, whilst on the bottom we expose how it affected the GDP and unemployment rate. Moreover, as before, we can find a slide on time working in the upper part of this screen, hence, we design a clear separation of those parts adding a soft blue mark.

- On the top left part, there is a **basic choropleth** chart where the user can visualize the change in the rate of mobility of groceries, that is the ratio of change of the frequency that people went to buy groceries at the shop. The metrics given are according to how far away it was from an expected baseline (before the pandemics). It is interesting to see the decline of this kind of mobility during the hardest part of the epidemic and how it increased after the worst part. It is in fact curious, to see that the tendency on moving to the grocery stores is higher after the pandemics than it was actually before, which makes us consider that the population has changed some conducts or daily habits due to the pandemics, moving ore frequently to the store presently. The colors chosen are a palette of colors that go from orange to blue. The user can see regions in orange when there is a decrease of mobility according to the baseline, and in blue where there is an increase.

- On the top right part, similar to the previous one, there is another **basic choropleth** chart. The idea is the same as for the previous one, but this time we show work mobility per regions and time. It is interesting to see how the two mobilities correlate, and we placed them one next to the other to make the correlations more accessible to the user. Note this time that there is a considerable decrease of mobility to workplace, clearly correlated to the lockdown we went through. However, the tendency of working from home has stayed after the pandemics, showing again how the disease affected the people's socioeconomic's environments. Colors are chosen with the same idea as before.

- On the bottom left part, there is a combination of a **line** chart with an **area** chart, following the same objective and procedure as aforementioned on the first screen. The line chart represents the evolution of new cases throughout all time according to the selected region. The area chart shows the GDP along the time, also by selected region. We decided to plot them together, since our goal is to find correlations between these variables. Notice that there are also two y-axis because there's no meaning on studying them in absolute value, and we focus on tendencies.

- On the bottom right part, there is also a combination of a **line** chart with an **area** chart. The line chart is the same as before, the evolution of new cases throughout all time according to the selected region. The area chart shows the unemployment along the time, also by selected region. As before, we decided to plot them together, since our goal is to find correlations between these variables.

Notice that there are also two y-axis because there's no meaning on studying them in absolute value, and we focus on tendencies.

With this second screen, the audience will be able to answer minor questions(in our point of view). For example, with the top right chart, the audience will be able to perceive how the illness affected the mobility of the population for every CCAA. With the top left chart, the audience will be able to percieve if it is more common than before work from home or not. Finally, wit the bottom charts, the audience will be able to answer economics issues, for example, how economy gets influenced by epidemiological issues. Additionally, on the one hand, they would be able to have a global and complete vision on how the pandemics affected directly to the work environment and employment rate, since we can see how workplace mobility decreased and the significant increase on the unemployment rate that matches monotonically the new cases distribution's behaviour. On the other hand, they can see how the impact in the population closest environments and economy, since we can observe the evolution over the pandemics time interval of the mobility to grocery and food stores (which are actually vital places people need to go to survive) and the sudden decrease of GDP (Gross Domestic Product) when the new cases of COVID-19 started to rise significantly.

3 Implementation

To finish the delivery, we have prepared the final dashboard with Tableau. The ideas mentioned before were implemented, and we tried our best to achieve the goals and answer the questions that the audience could have. The deliver consist in this pdf file, the Tableau source, the datasets used for the charts and also, the notebooks used to clean the data.