

AUTOMATIC TRANSLATION. ORIGINAL TRANSCRIPTION OF THE FOCUS GROUP IN SPANISH

----- STARTS THE FOCUS GROUP -----

Moderator1: We are going to show you a series of very simple visualization graphs on the COVID scope, and we are going to ask you some questions so that you can comment on what you see in these images. So if you want, we will start with the first of the images. And then we will move on to the first of the images that we would like you to comment on, and simply explain to us what you see in the image and what data you can extract from it.

*** IMAGE 1 ***

GFEXP_07: Well, the world is coming to an end. But apart from that, we can extract data, more or less an approximate idea, not very precise, of the number of cases per day; of the number of new cases per day over the last nine months. And also the evolution, a little bit the general idea of how the issue is, how it is evolving, and also a little bit the idea now that we are going with sawtooths. Someone might ask why, and well, we know why, don't we? Because the weekend data is not being accounted for, I guess, in the same way, or something like that. That's what I get out of it.

Moderator2: Any further input?

GFEXP_06: if you allow me, basically the scale is a bit problematic. You do realize, don't you? But, for example, as regards the lower part, which would be centered, we all know the background, so it helps us in the interpretation... But between May and July, for example, those, the oscillations that may occur are less, that is to say, they can be less differentiated in the graph. Probably, taking into account those sawtooths that GFEXP_07 also mentioned, although the trend can be seen, it would be much better to observe a trend graph. To understand for example that final part on the right of the curve where it drops abruptly. Well, I do not know if it is well or badly explained, if that corresponds to a specific data, right? These are things, more or less, that I see. Perhaps some numerical reference is also missing, as for example the vertical axis as one moves away to the right and if this were to increase the scale of data even more, it would make it much more difficult to observe whether we are talking about 12,000 or 16,000 for example, it may be important at a certain moment.

Moderator1: Any further input on this image?

GFEXP_05: about what GFEXP_06 said, the final part where the cases decline a lot (I understand that the cases), I don't know if it could give information that the pandemic has suddenly ended, that it has been vaccinated or that there is a missing data in the series and the graph has gone down... and then with respect to the sawtooths, I understand, as GFEXP_07 has said, that the steepest valleys are on the weekends, but I would be curious why there are peaks down, why there are valleys in something that might be the middle of the week. And it also looks good as, I understand again, this is an actual chart on the COVID and there's a little bit of bias with respect to what I'm thinking. But you could see the change in methodology in terms of the weekend - non-weekend sampling as the months progress. Weekend, not weekend or mid-week or whatever that is at the end.

Moderator2: perfect, any more input?

GFEXP_06: if you allow me, there is one that shocks me: it says daily new cases, it says date of daily data publication... at no time do we see what the source is either, right? The same as, for example, a legend at the top, a title of the graph... to know what we are looking at.

GFEXP_02: Yes, that is what I was going to say... new cases every day, of what? Because we don't really know if they are deaths or if they are contagions. Well, we know because of the context we are in, but we wouldn't know.

GFEXP_05: the location of the data, too.

GFEXP_01: that, that, I was going to comment the same thing. The location of the data, where are the data from, right? And we know the information that Moderator1 gave us: it is a COVID graph, so we assume that they are new cases diagnosed with COVID, but we don't know from what or by what method. It would be nice to know the method of diagnosis.

I am also very struck by the sawtooths, because we also see that there are very pronounced downward peaks, but we also see some upward peaks, so that should give us information of some kind of data collection, of how the frequency of data collection is being done, or is that captured in the graph in some way.

GFEXP_06: We could even add, let's say, in some specific point of the graph, some bar related to some specific event. For example, we are talking about COVID, we do not know at what moment is, hey, the alarm state is established, at what moment it is established in the de-escalation, the different phases, etcetera. That could also be interesting in order to be able to evaluate the graph a little better.

GFEXP_04: and also in this sense the change of methodology that GFEXP_05 commented before would also be good, i.e. when did it come into place.

GFEXP_03: it can be seen that, from August onwards, there has been a change in the data collection methodology that could have been used, for example, a one-week moving average or so, to smooth the line graph and see the trend a little better because what is clear is that it is difficult to infer between the months of August and November, what the average value of these new cases is. So it would have helped to understand the graph better, and that along with a notation in the visualization itself about when the data collection model was changed, as it seems to be the case, would have helped I think.

GFEXP_01: There is other information that I have also missed, that is, the percentage, the ratio of those diagnosed now, it has nothing to do with the percentage. I understand, because we know or we have a frame of reference, of those diagnosed in March. So in percentage, although the number, in percentage it may be very different.

GFEXP_06: even if you hurry me, maybe the left part of the time scale from January to February does not contribute excessively to know that there are zero cases, and it could be used, if the problem is of space, to stretch the series a little. Because right now from January to February it does not contribute absolutely nothing.

Moderator1: Are there any positive or negative aspects (you mentioned some) that you would like to highlight from this image?

Moderator2: in addition to the above.

GFEXP_05: the graphic is simple, it can be understood at a glance, even if the person is not very literate in this type of visualizations or graphics.

That is a good thing. And with respect to things like the scale, it's true that by making a transformation of the scale you could see in a better or more precise way the evolution. I don't know, if we use a logarithmic scale in certain parts of... if we make projections on different parts of the scale or so, but for an average reader it would be quite clear.

GFEXP_06: as a positive aspect, it gives a very clear impression to those who see it, knowing what it refers to, of what the trend is, that there is a very large valley there, and how it then begins to grow with those peaks.

GFEXP_07: Yes, and then it is true that the sawtooths are misleading and so on, because of the methodology used to collect data, but it reflects this methodology. In the end it reflects the data we are receiving on a daily basis, in a totally, well, reliable way.

GFEXP_01: I am going to talk about something negative. It seems to me that this information may be confusing. That is, contrary to what you are seeing, it seems to me that this is not giving us clear information on the situation, I think. I see cases here, but not having other information, such as the methodological part, such as daily data, etc., I think it may be giving information or a meaning of this graph, may not be reliable... well, it is reliable because it is supposed to be the data in numbers, but the semantics of this graph may be a little false or biased or biased.

GFEXP_03: I agree with that view and if I were to see this chart out there online, with all the things that have already been said... it would raise my suspicions about the reliability or the trustworthiness that I would place in this chart, in the information that I am seeing. Especially also because of the lack of communication of uncertainty, of what was discussed earlier about how many people were tested, on what date, you don't know what the percentages are, there's no communication of uncertainties whatsoever. And with all the missing legends in the data and so on... the reliability that I have in this graph is very low.

GFEXP_06: I would like to qualify what I said before just to make it clear: I agree with everything we have been saying. My previous comment was basically referring to what GFEXP_05 mentioned: that in itself the visualization is simple and it is quite clean.

Moderator1: Any more input on this image? Well, if you want to move on to the next one. I also tell you, they are all from the COVID area and so we also wanted to ask you what you observe here again, and what data you can extract.

*** IMAGE 2 ***

Moderator2: you can add, compare...

GFEXP_04: Well, the first thing is precisely what we mentioned before, that the part that used to look like a continuous line in the last few days, now we see bars, and I have many doubts as to why this is so. Because it is no longer a continuous line, but rather, I do not know if this is because the number of cases is decreasing every day in a very abrupt way, what is happening there. I have a lot of doubts.

GFEXP_06: I personally believe that all the comments we have made before in terms of captions, fonts, etc. apply. Simply referring to the visualization: the visualization seems to me to be less friendly for the observer, although it is true that, if we are talking about only one image, I think it is preferable almost

even the previous visualization. If we were talking about an interactive image, it is possible that by scrolling with a mouse over the graph it would be much easier to see the passage from one bar to another, because you can basically do it all along the whole path while in the previous case, supposedly, you would have to go directly to the line, right? That's the only advantage that I would see to such a representation over the previous one. But in general I think it's less friendly for the person who's looking at it and interpreting it.

GFEXP_07: yes, I agree. With everything.

GFEXP_01: I also agree with everything that was said, it is less user friendly than the previous one and applies everything that we have commented above.

GFEXP_02: I don't find it less friendly, because it seems to me that having padding gives me more the feeling that they are real cases, I don't know, that we are talking about things. In the line alone and with as many inputs and outputs as it seemed, well, I don't know... A personal appreciation, of course.

GFEXP_05: I'm with you there, GFEXP_02. If I wanted this graph to serve as an educator, or even to put fear into the real number of cases, the shading of the case area makes the visual impact greater from the beginning. It is true that in the bar/histogram part... it is a bit more complicated to read.

GFEXP_03: I think it supports different tasks, they are two different cases. Here you can see a little bit... perhaps the evolution of the cases over time is interpreted a little bit worse, but since it is a bar chart, in this case it allows you to compare contiguous days or the levels of new daily cases between two consecutive days, or even the same week. Then it would also allow you to analyze differences between months or days more distant in time, if you had some guidelines... some dotted lines coming out of each of the ticks on the two axes. Because in the new daily cases, if there were some horizontal lines, they would also allow you to make those comparisons a little bit better at a longer distance, right?

GFEXP_01: what I do... what you are commenting, what GFEXP_02 said, that they look like cases... I do see it with the accumulated cases, that is, to represent accumulated cases, it does look better, doesn't it, the number of accumulated cases when we see a fill, doesn't it? In the other case we see better the evolution over time of the cases, the peaks. But here the cumulative probably can be much more significant. We see that, even in the whole block from July to November, we see a large accumulated block.

GFEXP_02: sure, that shows the seriousness of the situation. More than just a single line. I see it.

GFEXP_06: to add something, I personally find it a bit shocking to have used only one color. This type of visualization supports the possibility of color grading or using different colors for different thresholds, which could also be interesting when it comes to being able to interpret the numbers behind each bar much better. I don't know to what extent it might saturate the person viewing it, but it might even be helpful.

GFEXP_07: well by changing at the saturation level instead of changing shades, yes it could be... it could be.

Moderator1: Are there any, again, any more positive or negative aspects that you would like to highlight from this image?

GFEXP_04: just a small comment here. Here for example what we have talked about previously about the change of scale, here it does not apply. That is, being cumulative as you are saying, with a logarithmic scale it would not make sense, or at least that is what I think. However, with the new daily cases, if they are from COVID, which we understand that they are, I do see it as you are commenting: that the magnitude is more visible in this way, on this scale.

GFEXP_02: in fact, what was said before the end of the figure, that what we did not know was appearing here does not happen, because as everything is filled in, we know that we are in that situation.

GFEXP_06: I don't agree with that, I still don't understand that part of the graph on the right.

GFEXP_05: it is too abrupt, yes, the change.

Moderator1: Okay, and anything else you want to add to this? Well, if you want to move on to the next image and nothing, same dynamics: comment a little bit about what you see, what conclusions you can draw from this image, and so on.

*** IMAGE 3 ***

GFEXP_06: that one sounds a bit familiar to me. Who's in? I'll start if you want. Okay, we know it's COVID, right? There's cumulative incidence rate, as of a particular date, cumulative incidence in... we have to over understand a lot of things.

Once again we have a lot of background behind, because we have oversaturation of information about it, and I think we can infer things from this graph that at first glance we would not have. I for example know that this is the map by districts of Madrid, but also because I am from Madrid. At least it seems so to me, because number 8 has to be Fuencarral-El Pardo, and there I see Puente de Vallecas, and so on. We do not really know where this map is from, right? Once we have internalized it, yes, we see some numbers that we do not know what they correspond to.

Then the color scale, well, you can get the idea... but the shades, let's say they are a bit... you have to go all the time to the legend and try to find the color to see what it corresponds to, and we don't know the exact number either. In other words, the difference between 150 and 450, because they are two different colors, but the variation is almost as much as the jump of two colors, right?

That way, aspects... Then you continue applying previous things: the font, the title... It is supposed... I do not know if this exercise on graphics that are static, or if we can suppose that behind there is some part of interactivity HTML, because there would already enter all the information to show when dragging the mouse for each one of the sites, I understand.

Moderator2: we cannot tell you.

GFEXP_06: because in principle, of course, all these graphs with supporting information behind them can improve a lot... They still have, they have many aspects that can be improved that I think are what we are looking at. We're going much more into what these charts don't have, rather than what they can have to improve them. Or what's behind them that we don't know. These are general impressions... then, if I think of some more as you contribute with the rest, I will tell you.

GFEXP_01: I think that, for example, I did not know where... what he was referring to. Because I am not from Madrid, and much less would I know which district it corresponds to. So, this information for a person who is not from Madrid, or is not familiar with the region, would not give him information. So, well, I do not assume anything, I am just looking at the chart and giving my opinion on what information

is missing in the chart, or in the legends, source, etcetera, that information would be missing there, wouldn't it, and at least where it is about, and where.

Then the issue of the choice of color palette. I understand that, well I don't know if that palette of gradients from blue to violet or maroon, because I guess they are trying to see the severity of the case, in terms of more intensity we have in that color palette... but as colors themselves, well they are not really conveying functionality to the user. A green: perfect, right? Colors that are semantically understandable, I mean, by the user. That we can transmit red is the most serious thing, isn't it, or play with the intensity of red and move in a question of transmitting color and semantics with that color.

GFEXP_07: I in this of the colors... could be an option, but another would be to use the same as before, to be the same hue with different degrees of saturation, it would probably end up giving a little bit the same information. Come on, the information we want in a clearer way.

I think that right now it is more or less clear, eh, in which cases there are more or less cases... that is, in which districts there are more or less cases. In other words, the choice of the palette is not disastrous, but it is true that there are some chromatic jumps that can be misinterpreted a little bit, because maybe there are differences of a jump that are visually interpreted as if they were bigger. I don't know if I'm making myself clear. So just by playing with one parameter, which would be saturation, it might be easier.

And nothing, it is also a little bit like what you have said... that you need to label to know which zone you are talking about and the number is obviously not enough to know which district we are talking about for people who are not familiar with map interpretation. For example, it happens to me in X, that I look at the map of X to see the COVID changes and I don't recognize my zone, because the distinction of the zones is very strange, then something else would be needed.

GFEXP_04: yes, in that sense, maybe, if you go that way, put a layer on the back that specifies you, that you can see, I don't know, Google Maps on the back or Google Streets or whatever, that lets you know that you are in Madrid and that you are not in X.

GFEXP_05: I am infinitely bothered by the scale of the data: cumulative incidence rate of what? 150 to 300 over what? Over the total population without taking into account the total that there is? Over every 100,000 cases as my bias learned in three months invites me to think? I don't know, if we pull, well, if I didn't know it was Madrid, the map is super strange. Especially, in relation to that, with the scale of the data. And if you already know it's Madrid and you bring all the biases from home that you can, you could even define the kind of socioeconomic situation of each of the health zones that are represented there, or the districts about or whatever it represents itself.

I understand that the district number has to do with a classification of some kind, which should be nearby, in a legend on the map, saying which neighborhood it is, which district, which health area, whatever... for people who are not from Madrid, knowing that 1 is I don't know which region, or I don't know which area, would be quite helpful.

It would be good to know what the cumulative incidence is. Not only in terms of scale, but what is considered, for how long it is measured, if it is an average over a day, over a week, over a month?

GFEXP_06: you are right, about the period covered.

GFEXP_05: sure, it is reported as of day 15, but you don't know that calculation over what time period it is made.

GFEXP_07: yes, and well, apart from the headline, although we are supposed to be experts in this type of things... it is possible to make the typical link, to make a complete description of the calculation behind it, which I miss in what I am seeing out there.

GFEXP_06: I'm seeing one very positive thing, now that I'm looking at the map a little bit, and that is that I think almost all the colors of the scale at least have representation on the map. I've seemed to count, except for the less than, not the less than 150 which I'm not sure if it's neighborhood 7 or not... that's maybe the only problem I have, I think the rest of the colors are all represented.

GFEXP_05: about that... about the one of less than 150 I have doubts whether it is the 7 or the empty one around, for example. That also ties in with what you say about the scale probably not being accessible, the color scale used.

GFEXP_03: I am going to say, somewhat contrary to the general opinion, that the scale does not seem bad to me, the choice of colors. The only thing is that there is that problem that GFEXP_05 and GFEXP_06 just mentioned, that it is difficult to know if zone 7 really corresponds to less than 150, and also the others that are around, to solve that, a more marked line could have been made around the zone that is being represented. That is, around all the zones of the map.

But I believe that the scale does meet the objective of, at a glance, or almost in a pre-attentive way, it is clear that zone 13 is the worst, the worst, come on, in terms of the incidence rate, which should also have been explained. But it is seen, for example, in zone 7, as an island in the middle of the map, that one would wonder why that happens and I think that this is enabled by the color scale, in that sense, it is going well. The only thing, the detail of making a more marked line around the zones, for me would be.

GFEXP_06: I think GFEXP_07 hit the nail on the head there, with it not being the best possible, but not bad either, in terms of palette choice.

GFEXP_02: yes, I agree too. It doesn't look bad to me and it's easy to visualize. If we pass it black and white, too.

GFEXP_01: and here talking about moving it to black and white... the accessibility issue with the color issue.

GFEXP_06: I was thinking about that, what happens is that I don't know how a person who has problems seeing this type of colors sees. I don't know what kind of person, what kind of visual impairment could make them not see correctly with this range....

GFEXP_03: I think this is "color-blind safe", this scale. And it sounds like Color Brewer to me, too.

GFEXP_02: I don't think it would be much of a problem, I don't know.

Moderator1: so, any other aspect that you would like to highlight from this data or the image shown?

Moderator2: if not, we jump to the next thing.

GFEXP_05: I would again emphasize the numerical scale. It drives me crazy. If you take it with a grain of salt, region 8, as big as it is, is either unpopulated, or it is the safest region where you can go to live, even if it is not the lightest color of all.

GFEXP_04: Yes, well in that sense also, yes, to put a little bit, what do I know, what you were commenting: if it is about one hundred thousand inhabitants, how the calculation has been made, and

in this sense, whether or not there is uncertainty, which I understand there is and I do not know, it is not mentioned. But well, that could be solved with an interactive graph. And that's it.

Moderator1: okay, so if you want to move on to the next image, and the same thing: what do you see in the image and what would you highlight.

*** IMAGE 4 * // It is the improved image 1.**

GFEXP_06: I see that you have included some of the recommendations previously mentioned. Yes, we had talked about events... maybe it is not the best way to put an event and especially when we are talking about, not only interactively, which would surely be done in another way, but it seems very strange to me the location of, right, where that legend is.

GFEXP_03: for me, this chart is even less reliable than the previous one, precisely because of that. It gives me even more of a bad feeling.

GFEXP_01: First of all, I don't know to which region this corresponds, but of course, here in the Canary Islands neither rapid nor antigens, nor probables were counted before, and I know this first hand, so... if I don't even know the region it seems quite confusing to me, very confusing.

GFEXP_07: but that, there we are already criticizing the data source.

GFEXP_01: so, I don't have... not having clarity, if I don't have reliability in the source of the data, I mean, I don't know where it comes from and I don't know the source of the data... One main thing to be able to believe a graph is to know the source.

GFEXP_07: Yes, but I don't know if maybe Moderator1 and Moderator2 can tell us if... because, of course, in the end we know that anything that has to do with COVID data in Spain... anyway. We are going to... right? In other words, there are many problems with the way the data have been collected over the months. But well, going back to the visualization, it would be easier, or more intuitive, something that would show directly where the border is and that, perhaps, by putting a line where the border between the two is... at the moment when the data collection methodology is changed, an arrow or something to the left to show what happened until May 10, and something else to the right to show what happens from May onwards. I would see it as more intuitive, easier to interpret.

GFEXP_02: yes, I agree with that. Also, that it doesn't give much more information to have the two colors if we already have the legend that is telling us until May 10. Come on, for me it is not much.

GFEXP_06: if probably just what it says what GFEXP_07 says: a vertical bar there, with a little dot where it then explains what it is. In fact, the colors would not be necessary and because we are already very used to the PCR terminology, okay? I don't know about ELISA, maybe I'm missing something, but I can't say directly what ELISA diagnosed refers to.

GFEXP_05: I looked it up and it's the one for... the antibody one.

GFEXP_06: no, but it simply adds confusion, it puts in the graph a text that can be solved in a thousand much simpler ways. And simply the vertical bar without color change may be enough.

GFEXP_02: much clearer bar.

Moderator1: any other highlights from this?

GFEXP_06: because it has some additional information with respect to the previous one. Maybe not put in the best way, but it gives some additional information that allows... I would wonder if it allows to understand something, okay? Because we see that when there is the change there is a similar trend, the only thing we could say is "hey, is the part on the right higher because of that change in diagnosis or because there are just more contagions?"

GFEXP_07: maybe, again here the questions that arise. Because one says, it turns out that until May 10 we included more data apparently and there are fewer cases. So well, there are fewer cases overall, in the area under the curve, so maybe this requires, well, a link to a site that explains a little bit all the background information.

GFEXP_06: Basically, I think we have become so used to digitally consulting interactive graphics that these things fall a little short. I mean, if we see it in a printed newspaper I don't know to what extent we would think that they are not offering relevant information.

GFEXP_03: I have a very bad feeling about this chart. By now I would have thought that we would have gotten a callus with what this is.

Moderator1: but in what way?

GFEXP_03: I have a bad feeling that now they are putting this information and that it is not visible... I mean, I would expect the change of methodology to have come in July, right, when there were those picks, that saw... but it is just something I did not expect. So, that extra information, instead of saying to myself "oh, that's good, they have put in extra information", what makes me think is "oh, they are trying to deceive me by putting in probably biased information"... or something that is not... something is escaping me. So no, at this point I would decide to pass on this chart and look at something else, I don't know.

GFEXP_06: I fundamentally believe that the feeling may be that there is a lack of additional information to help us interpret all the data much better if we want to take a more critical and personal view of the information we are consuming.

GFEXP_05: with what GFEXP_06 said before about ELISA, the use of vocabulary is important because I do not know if anyone here knew what the ELISA method was, but surely if they had put antibodies, which is what the media mostly use, or what we experience even in government press conferences, we would have found out what type of test was applied depending on the time of year.

Moderator1: Anything else on this? So let's move on to the next question that has to do with this. You can't see them, but they're the ones that you saw at the beginning and now... And simply, if you had to illustrate a news item about, well then about the COVID situation in Spain about this data, which version or versions of these that we've shown you would you choose?

GFEXP_03: Is it mandatory to choose one?

Moderator2: yes, or if you have buts... or none.

GFEXP_03: but my question is: do I have to choose the one I think is better or can I say I don't choose.

Moderator2: justify your choice and that's it, you can say no and justify.

GFEXP_03: Well, no, I would not choose any of them. Because there is a lack of data, a lack of information, I would not choose any of them.

GFEXP_02: well, if I had to choose, I would choose the one below... it has more information.

GFEXP_05: if I had to choose, I would choose the one above, even though it has less information, because what it adds seems unclear to me. And I would delete the brutal drop in November because I think it is misleading. And if you add where the cases are from or something else like that... if it is from Spain, if it is from Alpedrete below, wherever... it would be great.

GFEXP_04: yes, I agree with GFEXP_05. If I had to choose between the two I would choose the one above, and it is more what GFEXP_03 said before, that I agree with him, that I do not see any perceptible change there and I know that this change exists and that it can be well visualized and approached in this period of time, because you can clearly see the change in methodology. But here it does not bring me anything, in any case if you put, as you all said, a vertical line but I do not know. Obviously I'm with GFEXP_03 as well, and would prefer not to use either, but if I had to choose the one above it seems a bit more sensible.

GFEXP_03: I would not choose any of them, but if I were forced to choose, I would choose the one above.

GFEXP_01: me too, the one above, but taking away the sharp drop in November...

GFEXP_06: I am not very clear. I, with all that has been said before, clearly I would not choose either one of them. And I have to contradict myself... I think I like it better if there is additional information, right? In that case I would say the bottom one. Eero is that I don't like the bottom one... I don't like it at all. I have a problem with that, in principle I would prefer the bottom one for more information, but... although the information does not add too much, but it is very poorly presented so... I don't even know if in my opinion, not so much for me because I don't like either of the two as for other people, I would rather leave the top one because, why add more mess? I just don't like it.

GFEXP_07: me too, I think the same as GFEXP_06. In the same way I want to say that in general I like the fact that there is more information, but it is true that the information that is added to the one below opens more questions than it answers, probably. So, well, we can leave the one above and leave that general view of the number of cases detected, which I do know... Maybe by changing that word, it would help a little bit... it would be more honest.

Moderator1: Anything else to highlight from this? Okay so we'll move on to the next image, which is this one. And again what do you observe and what would you highlight from this new image.

*** IMAGE 5 * // It is the improved image 2.**

GFEXP_07: very good, there are the weekends! Well, it helps to understand one of the questions that was asked with the graphs, right? What is happening with the data collection directly. That is, here more than information on the evolution of the detected data, we could use it to question how the data is being reported, but well, it does help a little bit.

GFEXP_06: I think... the same applies as I said about the location of the legend and so on, but evidently that information is "somewhat" better displayed, and as GFEXP_07 says, it allows to give some additional explanation to the previous data.

GFEXP_03: I agree, this one gives me a better feeling.

GFEXP_02: me too, clearer.

GFEXP_07: and in this one even the legend is not so clear that it is necessary to change the place, although it is possible to put there the little arrow or the vertical line... but well, here it is not the same thing.

GFEXP_06: I mean in terms of location I would always expect to have the legend, I don't know, to the right, outside the graph; not there on top, but yes, in terms of what the legend says, it is much better understood.

GFEXP_05: it seems to me too long the legend of the "from July 4 the data update occurs on Monday, the day on which the cumulative incidence data from Friday to Monday are presented". Uhm... weekends as of July 4 partial daily update.

GFEXP_07: yes but... Mmmmm... because more than a legend, which is true, it is an explanation, it is a notation that... well, it is fine.

GFEXP_05: eh sure, but why include all that information in the legend even though we are, I think we all agree, that the more information is well displayed, the better, in this type of graphs and not... that: a short legend with an asterisk that takes you to the reporting methodology.

GFEXP_07: ah, that's right.

GFEXP_06: I see one thing that catches my attention, we assume that these data are correct. I have not followed up the daily data, but the peaks observed there on Fridays, when one looks at the following peak on Monday, which in theory is for the whole weekend, that is, accumulated incidence from Friday to Monday, and yet it is lower than usual, I see in the series, than that of Friday, and then it grows again. It would give food for thought. What happens, is it that on weekends it is not that data are not accumulated, but that tests are not done either, or less tests are done...? It strikes me when I see it in the graph, it is a question I would ask myself.

What happens that on the weekends the cases do not grow when we are reporting on Monday cases for the whole weekend, and yet it is a lower value than on Friday and from then on it goes up again on Monday, Tuesday, Wednesday...? It only calls my attention.

GFEXP_02: and I think the same thing, it's very rare...

GFEXP_04: yes, I also agree that you don't know if some cases are being extrapolated, why on Sunday there are fewer than on Saturday, why that is a general trend.... It doesn't make that clear to you, so an asterisk that leads you to how the methodology has been done seems to me... it seems right to me.

GFEXP_03: still the one-week moving average would be very good... Here a red line above.

GFEXP_04: totally agree.

GFEXP_06: yes.

Moderator1: Okay, any more of this? Ok, so we ask you the same question as in the previous case: if you had to illustrate a news item with allusion to this data, which version or versions would you choose?

GFEXP_07: the one below.

GFEXP_02: I the one below, too.

GFEXP_06: I have no doubts in this case either.

GFEXP_05: the one below.

GFEXP_06: the one below, of course.

GFEXP_04: the one below, yes.

Moderator1: why in particular? by what aspect?

GFEXP_02: because it has more information and you are not left wondering what happens in that weekend change.

GFEXP_06: basically because the information that is added allows a better explanation of these changes, that is, I think it is useful information for the person who is watching it.

GFEXP_07: yes, at least there is an anomaly of the one in the graph that is explained a little better with the colors.

GFEXP_03: I agree.

GFEXP_04: yes, in contrast to the other two that you gave us to choose from, here the extra information does contribute. That is, it gives you information that is interesting and that you can interpret.

GFEXP_03: explains the weirdness of the previous one... the surprising fact of that second part of the graph as of July now you say "ah well, it will be because of this". It still has a lot of problems.

Moderator1: Okay, so if you want to move on to the next image if you don't have anything else to contribute to this part. And it's the next one. And the same thing, what do you see now, what data do you see now and what do you take from this image.

Moderator2: it's not "one and one", it's both, together.

Moderator1: yes, we consider this image as one.

*** IMAGE 6 * // It is the improved image 3.**

GFEXP_07: phew! I would change the color scale on the right to think... to know that we are talking about different things. But problems... I don't know.

GFEXP_06: it doesn't do anything for me. I'm going to have to go from one to the other, and I'm going to check the scale all the time without knowing exactly how I'm comparing. I am going to look at a district, based on the fact that they are the districts of Madrid, and I am going to look from one to another trying to compare... I cannot establish comparisons well. In other words, when it comes to care, simply because of the phenomenon of care, I am going to have to review them almost independently and try to draw some global conclusions that I can then more or less integrate into my corpus. Yes there... yes you can see that parts 11, 12, 13 and 17, which are clear in one place, are obscure in another. But the nuances of the information can escape. In other words, when it comes to relativizing one with respect to the other, I see it as much more complicated.

GFEXP_03: I agree, drawing general conclusions from the map is complicated, because of that, because of having to go comparing and moving the eyes from left to right and vice versa to establish comparisons. Surely what people will do when they see this will be to go to the map on the left, which is the one they saw before, and identify those regions that are especially interesting, that is, the clearest, 7, or the most saturated, or the one that is higher on the scale, 13, and go and look for its counterpart on the map on the right to see if a relationship can be established. To see if the one that has less is the one that has more in the other, and the other way around, or to see if we can infer something from that comparison. But it is not possible, with this representation, to draw a... to extrapolate something in a general way from what is happening in Madrid... difficult.

GFEXP_07: man, yes... I mean, let's see, this graph confirms something that you had, let's say, suggested before almost... for almost all the areas. And it is true that I think that is what it is, that it has a high cognitive load, the interpretation because it forces to go, as you said, from left to right.

GFEXP_02: I see a problem with using the same color palette... they are different things. And it is difficult to compare things, believing that we are in the same... measuring the same thing. I would never use the same scale, that is, the same color.

GFEXP_06: even if what you want to see is the relationship between cumulative incidence rate and average income per household, which by the way the dates are three years apart, but let's assume that the average income does not change that much... why not combine it in a graph if that is the information you are interested in? I don't know, it depends on what information you want to provide to the... I'm going to say "reader", but yes, to the person who is visualizing the graph. If what you want to do is to show the relationship between incidence rate and average income, then why not show the map directly. Then you can provide additional information, as we have said, in a digital medium, or even put the cumulative incidence rate, if you are interested in this information, and then, mixed, the incidence rate by income. Otherwise, if it is established to make a comparison, I do not see the sense of making two graphs.

GFEXP_03: that's why this chart is also suspicious. Because in fact it happens, I at least what I have analyzed in this graph, while as GFEXP_06 says a bi-variate map could have been made, to show that... if I go to the region... The dangerous thing about this is that I can infer things, as I have explained, at a general level, that only happen at a particular level. For example, region 7 on the map on the left is the one with the lowest incidence, cumulative incidence rate, if I go to the graph on the right it's one of the most... of those with the highest average income. If I go to 13 in the map on the left, it is the one with the highest incidence rate... it is the other way around in the map on the right, the one with the average income. Perhaps simply by establishing these two comparisons because they are the easiest, I can extrapolate that this is the case throughout the community of Madrid, when I would not be sure, so... It may be a little suspicious, or perhaps I can interpret that the people who have made this map are trying to make me reach that conclusion, if I get a little...

GFEXP_06: what Alex says makes a lot of sense. If we then add to this seeing 2, 3, 4, 5... which have the same cumulative incidence rate, and yet all have different average income, in the end the interpretation can lead to a lot of bias depending on which psychological anchor is being used by the person viewing the information.

GFEXP_05: totally, if I look at zone 7 on the left and I go to the right, at least the first thing I try to look for is a similar zone, and I go to 13 continuously. On top of that, with the little contrast that the numbers have in the 13, I am not able to see anything else at the first glance when moving from one graph to another.

GFEXP_04: Yes, I was going to comment that it makes me a little dizzy. I mean, I don't know, none of you have said it, but it makes me dizzy. I mean, so much information even gives me a little bit of... it's too much and I can't assimilate what... what is there. And yes, it's true, there a little bit of correlation there doesn't have to imply causation, but there may be something like what GFEXP_03 says, that may be being a little bit manipulated so that I'm trying to see that.

GFEXP_06: in fact, the shading of the numbers I think does not help us, because when jumping from one to another with the color contrasts and having the number in white with a shading at the end it does get a little crazy.

Moderator1: Is there any other aspect to highlight from this? Well, same question as before, if you had to illustrate, again, a news item... which of these two images would you choose?

Moderator2: a data-related news item.

GFEXP_02: I see the top one more clearly. The bottom one makes me dizzy, as you said.

GFEXP_06: I would almost say that it depends on the medium you are, and where you are looking for. But honestly, the one above, the information within the problems that we have seen before of information that you have, is, let's say, more clear and objective. It would probably be the one I would use. If I am interested, when giving a news item, in trying to see that relationship between incidence rate and average income, I would try to do it in another way. I would never choose this representation, the one below, I mean.

GFEXP_03: the same, it depends on the headline you want...

GFEXP_04: yes, me the top one too, because it basically makes me less dizzy.

GFEXP_06: is a good reason.

GFEXP_05: I would choose the one above by adding, I don't know, an emoticon of banknotes by zones on a scale of 1 to 5 that will mark you when you are short and that's it.

GFEXP_02: that is very interesting.

GFEXP_07: yes, the one above as well.

Moderator1: Okay then, if you have nothing more to add, we are now going to ask you a final question about all these visualizations that we have been showing. Given that the data source is

reliable, which, in this case, it is... Do you consider that the use of any of these visualizations could be misleading?

GFEXP_03: Well, we have been commenting on the fact that there is information that is introduced, that does not explain what you are seeing, because it gives rise to suspicion on the part of the person who sees it. And the fact that they make you... they force you to establish a comparison, when it is not really necessary and it could have been solved for example with the emoticons of the banknotes, as GFEXP_05 has said, or with a bi-variate map, it makes you think. Then the amount of information that is a little "missing" from the graphs, the legends, the lack of the origin of the data, the methodology... has problems, of course. And I would consider that, as I have already said, that whoever created these graphs, I can understand that he is trying to induce an idea in me or to make me think in a particular way, no doubt.

GFEXP_06: If I don't see an adequate explanation, if I am just presented with these graphs in an article, I would first look at them and try to see what I can infer. I think I would come to the conclusion that, without additional information, well I can either dismiss them or try to contrast them. I would hope that the supporting text would lead me to be able to understand them. And even then, I might be suspicious.

There are aspects that, because we are familiar with them, what we have said on weekends, etc., among all the images, probably the one at the bottom right is the one that I would trust the least in principle, and I would try to look for other conclusions, go to the data, etc. So, it seems to me that it can give much more room for deception or manipulation, although in general all of them, because of what we have been saying.

GFEXP_07: yes, I, let's see, manipulation I would see the one on the bottom right, or attempt to make me think something that may or may not be real, okay? In the others, what I would see is always a need for more information, I would always be left with questions. Okay, these are the cases, but how many, what GFEXP_01 said before, how many tests have been done, or percentage, I don't know, other type of information that would give me more... other type of data or visualizations, that would give me more information about... well, how the pandemic is evolving in this case, right? But well, I think that the one on the right is manipulative and the others are a little bit lacking in information, and that's it.

GFEXP_06: yes, that, not that they manipulate, but they are incomplete, they would probably force you to look for more information.

GFEXP_02: I agree, the same.

GFEXP_05: I would love to read the article that goes around each of the images to see which way they are going to throw each of the visualizations.

GFEXP_04: maybe the article is "bad visualizations".

GFEXP_07: don't overdo it either.

GFEXP_04: no man, just kidding, just kidding. But I don't know, no, partly because of the lack of information I would also say yes, that perhaps, I never want to think badly, that there is someone who wants to manipulate me. Surely the one at the bottom right is the one that may raise more suspicions. The others may not be the best method to choose, nobody is perfect, I for example am now doing a similar one with epidemiology in X, and I am looking black to represent everything I want, and I know it is complicated and difficult. So, no, I don't want to kill the person who makes the graphs because he

can also make mistakes, obviously, he can have his faults. But yes, if it is a medium from which we are reading and we have or want to have some credibility, then I would try to add as much extra information as possible. The thing is that in the end it is the same as always, we have two dimensions to illustrate perhaps too many things and it is very complicated. That's why some people do theses on it, don't they?

GFEXP_07: yes, I do think a good labeling, yes I think it can be easy to add. I mean, someone was talking about the coefficient of how much we could rely on data and things... There are technical details that for the general public could be too much information, so of course, in the end what you show has to be simple so that it can be understood. But of course, the border between simplicity and the absence of data is not easy to know where it is.

GFEXP_03: it is very interesting, it is something that I have been thinking about for a long time, isn't it, the simplicity of the graphs and then accompanying them with texts that explain what the graph means. As I understand it, in order to be good, a chart has to be self-explanatory, so if we always have to go to third parties to constantly validate it, then why do we make the chart? And then, I have also adopted, personally, the position in which... I mean, I don't know when someone does it for that reason, for simplicity or because they don't know or because they might make a mistake... Or when they really use that as an excuse to manipulate me. So sometimes it is complicated, that border is also complicated to detect, as an observer who wants to be informed. So, I understand that it is a bit of a personal position, "well, my level of paranoia now I put it here or here, and I trust, I don't trust, and let's see in which media this is published", and many more... much more context that well, that you can evaluate, but I think we live in complicated times to trust everything a little, so I am a little more... but well, it's personal.

GFEXP_06: I personally, regardless of the accompanying text, believe that I could trust a medium in which, referring to the image in the center, the image at the bottom would be more reliable than the one at the top, for example.

GFEXP_07: yes.

Moderator1: And why exactly?

GFEXP_06: because it is giving me additional information that may be of importance when interpreting the graph. The problem, I think we have a problem, because, for that reason, everything we already know and the oversaturation of information we have about it, but if I see the graph in two media I will probably say "oh well, this one explains it better". Then, independently of the accompanying text, which may be more or less biased, justifying some actions or others... But it is giving me that additional information that allows me to explain why there are sudden drops, why it has that sawtooth aspect...

Moderator1: I would like to take this opportunity to ask one last question: what do you consider to be a visualization or a misleading graphic?

Moderator2: if you want to add something as you were saying, such as the manipulation commented by GFEXP_03.

GFEXP_05: introduce bad practices in bad faith.

GFEXP_04: yes, especially when you see, for example, I don't know, these graphs that we often see that you see a bar that says 100 and a bar next to it that says 800 and maybe the difference between 800

and 100 is nothing... and you say please, here you are trying to deceive me, you are doing it on purpose. Here in these I cannot see bad faith in any of them, or at least I understand... I am very well thought out and I think that no, that the person who has done it does not have bad faith, he is not... maybe the only one, the one that, I repeat, the one of the two maps together, but apart from that... I do not see bad faith on the part of the person who has done it.

GFEXP_03: I simply believe that in the end it doesn't matter a little, come on, the good or bad faith of the person who makes the graph is indifferent. If the graph is confusing at the end, it doesn't matter if he did it on purpose or not. So, for me it is indifferent. I consider that a graph is well done or not, then, if it has been done on purpose, then it is bad, because I should consider his praxis in general, and his moral scale, and then if it has not been so, then I should try to do it better.

GFEXP_04: yes, or at least rectify, say: "I made a mistake".

GFEXP_03: "it will not happen again".

GFEXP_04: that's right.

GFEXP_07: well, I don't know, there are graphs that try to tell stories, and without misleading you can draw conclusions. And an example of this can be the one at the bottom right, that is, you can... I mean, maybe a little more embellished, even if it is only with a footer or whatever, you can deduce a specific conclusion of a specific moment in the evolution of the pandemic. And it may be that the graphics help to express this information if we are talking more about a journalistic issue, not so much a scientific one, which I think are a bit... So, well, the use of graphics to tell a story, often makes the graphics be presented in a certain way, I don't know if that is misleading, but well, I don't know.

GFEXP_06: I believe that a misleading or manipulative visualization is one that omits essential information for the person viewing it to be able to interpret it correctly. In that sense, probably none of these six graphs in itself is manipulative or misleading, as such. There are some that can be used in a more manipulative way, already through the word, whether they are relevant or not. But when essential information is omitted, especially when it is accompanied by some explanatory text of the author's interpretation, near the graphic... is where the deception can come. But even so, a visualization can be misleading when, by omitting information that is essential for the interpretation, it can lead the person viewing the graph to reach incorrect conclusions.

GFEXP_07: Of course, in this sense, in these specific graphs, in those on the left, there is something missing that says the volume of tests that were being done at each moment, or the lack of knowledge about it, I don't know, maybe it is because we don't have that data. But now that we know everything we know, the interpretation here is clear, that is, there were many, many more cases, there are many, many, many more cases now in this second wave than in the first, but it seems that this can be questioned, right? So, is the diagnostic capacity we had in the first wave the same as it is now? There... there may be a little bit of bias in the interpretation that requires us to be given some data, at least to be able to doubt the interpretation. If not, in order to be able to interpret it in an accurate way, if not... if we do not have the data, at least say "look, here is this problem, you can do whatever you want to do with your head when interpreting the data". I don't know if I'm making myself clear, but okay.

----- ENDS THE FOCUS GROUP -----