Youth and COVID-19 data visualisation

Words: 9603

Anke Morsing (s2102692)

Supervisor: P.J. Klok & Guus Dix

Management Society and Technology

University of Twente
Enschede, The Netherlands

March 2021

Index

1. Abstract

In this thesis the research question *How do different factors influence the way youth comprehends graphs regarding covid-19 cases in the Netherlands?* will be explored. The factors that are explored are visual characteristics of graphs, the viewer's knowledge about graphs, and the viewer's expectations about the content in the graph and the background characteristics of respondents, gender, age and level of education. Before the research started the following hypotheses were conducted: that visual aspects of graphs, knowledge viewers have about reading graphs and level of education have an effect on the comprehension of viewers of the graphs. For the expectations viewers have about graphs about covid-19 cases, it was expected that viewers that are more worried about covid-19 and see the pandemic as a problem to comprehend the graphs different than a viewer that is less worried about covid-19. No relation of gender and age (within the group youth) on the comprehension of graphs was expected. For the research 10 interviews were conducted. The visual characteristics of graphs, the viewer's knowledge about graphs, and the viewer's knowledge and expectations about the content in the graph seem to influence the graph comprehension of youth regarding graphs about covid-19.

2. Introduction

On February 27th 2020 the first covid-19 case was discovered in the Netherlands. Since then the amount of covid-19 cases have fluctuated and so have different guidelines to try to keep the number of cases as low as possible. Just because there are guidelines to keep the number of cases low, does not mean that people are actually following these guidelines. In the press conference on the 6th of August 2020, Prime minister Mark Rutte said that more and more people do not follow the guidelines and that it was visible in the recent outbreaks. (Rijksoverheid, 2020)

Several variables influence why people decide to follow or not follow the guidelines, such as entitlement (Zitek & Schlund, 2021), social influence (Tunçgenç, El Zein, Sulik, Newson, Zhao, Dezecache, Deroy, 2021) and graphs (Romano, Sotis, Dominioni, Guidi, 2020). Romano et al. found that even though underlying data in the graphs are identical, that merely changing the scale can alter the worry, opinion of public policy and intent to follow precautionary measures regarding covid-19.

I think that covid-19 gives us a unique situation in that more people are looking at casual data visualisation than before. Casual data visualisation meaning data visualized to "*depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations."* People might have come across causal graphs on occasion, but during the pandemic, data visualisation has been used more than ever before by politicians and civil servants (Mathieson, 2021) but also by random Twitter users. (Trajkova, Alhakamy, Cafaro, Vedak, Mallappa, Kankara, 2020). This uniqueness of the current situation is where the scientific relevance lies.

In the same press conference as mentioned before, the prime minister, Mark Rutte, specifically mentioned youth when it came to not following guidelines. Warning them that if too many of them flout the rules they'll endanger their parents, grandparents and everyone around them with fragile health and it can make them very sick themselves as well. Rutte also appeals to the youth that has not been following guidelines to stay away from people with fragile health, since that is a risk that you cannot take. Later in the same press conference, Rutte says that the number of young people that do not follow guidelines is higher than the average. (Rijksoverheid, 2020)

Entitlement of people, and the social circle they interact with is not something you can externally influence, but the way graphs are represented to people is something that can be easily changed.

Because the number of young people that do not follow guidelines is higher than the average, this group will be interesting to study. Important in the support of the guidelines is the data visualization in graphs about covid-19, but which factors influence the comprehension of youth about graphs regarding covid-19 cases has not been widely researched. While Romano et al. (2020) studied the influence of the scale of graphs on the understanding, attitudes and policy preferences, this study does not focus on an age group specifically. Shah & Hoeffner state that one of the factors that influence the comprehension of graphs is the viewer's knowledge about graphs. Young people are often still in the process of learning about graphs and the role of education can be better studied when looking at a group of young people. To study this the following research question has been conducted.

How do different factors influence the way youth comprehends graphs regarding covid-19 cases in the Netherlands?

Three different factors will be looked at: visual characteristics of graphs, the viewer's knowledge about graphs, and the viewer's knowledge and expectations about the content in the graph. Additionally, the background characteristics of respondents, gender, age (in the group youth) and level of education will be looked at to see if they have any influence on comprehension.

Why these variables have been chosen to explore will be explained in the theory section. Two sub-questions have been conducted to help answer the main research question.

- How are graphs about covid-19 cases presented in the Netherlands?
- How do visual characteristics, the viewer's knowledge about graphs, the viewer's knowledge and expectations about the content in the graph, and background characteristics of viewers influence the comprehension of graphs about covid-19?

3. Theory

Data visualization about covid-19 is often casual, "*the use of computer-mediated tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations*" both in the way people come across data visualization and that a lot of data visualization is made by "amateurs" on social media. Analysis data is no longer restricted to research facilities where researchers are at work. Instead, now everyone has the possibility to analyse, study and interpret data and visualize it, meaning that a majority of people who come across data or data visualization are not trained in data analysis. The ability to effectively interpret data visualizations is particularly crucial right now because people rely on their analysis to make decisions regarding the behaviour in regards to covid-19 (Trajkova, et al., 2020).

Graph comprehension includes translation, interpretation and extrapolation/interpolations. Translating a graph requires the viewer to change the form of communication. For example by describing the graph in words or commenting on the structure of the graph. Interpretation asks viewers to rearrange material and sorting the important data from the less important data. To do this viewers can look for a relationship in a graph. Extrapolation/interpolation extends on the interpretation of graphs and not only looks at what the graph is depicting but also possible consequences by looking for trends or implications. Reading the graph shows the least trouble for students. When answering interpretation questions students tend to make mistakes. Extrapolation gives the biggest struggle for students (Friel, Curcio & Bright, 2001).

Three main factors influence the comprehension of graphs; visual characteristics, the viewer's knowledge about graphs, and the viewer's knowledge and expectations about the content in the graph. Especially youth does not have a lot of knowledge about graphs (Shah & Hoeffner, 2002).

Several steps go into comprehending a graph. The first step in reading a graph is translating the visual aspects. The way the graph is presented influences the ability to read the graph. Readers see continuous relationship/x-y trends (when x is increasing, y will decrease) easier in line graphs than bar graphs. But also data not continuous plotted in a line graph can make viewers see a continuous relationship when it is not there. This is one example of how which type of graph chosen can influence interpretation. X-y trends are best depicted in line graphs, bar graphs are good for showing discrete comparisons and pie charts for relative proportions (Shah & Hoeffner, 2002).

Other visual aspects of the graph also design influence the comprehension of graphs, the most common are colour, size and aspect ratio. Colour is used in graphs in several different ways. For metric information with levels of saturation or to differentiate variables from each other. This has both benefits and drawbacks. For example, colour can be mistaken as higher or lower than other colours, but colours make it easier for viewers to keep track of the different variables (Shah & Hoeffner, 2002). Which way colours are interpreted differ per viewer; different cultures have a different meaning for colours and so do different professions (Brockmann, 1991). Young people are less likely to have been influenced by professions in the way they interpret colours. The size of a graph and aspect ratio also influences the interpretation of graph viewers. Viewers see a higher correlation in a graph that is densely packed. Density can be increased or decreased by adding data points or changing the size of the graph (Shah & Hoeffner, 2002).

The knowledge viewers have about graphs influences the memory that viewers have of the data visualized (Shah & Hoeffner, 2002). Knowledge also influences graph comprehension, analysing data

is closely related to mathematics and many high school students have difficulty comprehending basic proportional concepts, such as per cent or ratio, and applying them to numerical data presented in statistical contexts (Gal, 1993, p.199) (Friel, Curcio, Bright, 2001).

Research has also shown that errors are made by viewers when interpreting graphs when prior knowledge and expectations the viewers hold is inconsistent with the data visualized in the graph. On the other hand, viewers whose beliefs align with the new knowledge and expectations are more likely to positively interpret the information. When viewers believe strongly that two variables are related they will overestimate the correlation between the two variables (Shah & Hoeffner, 2002). The expectations viewers have might be something the viewer will look to confirm. When a graph is complex the expectations are easier to see than the data visualized. Being familiar with the content of a graph also makes it easier to keep track of the information in a graph helping with comprehension (Shah & Freedman, 2009). In the case of covid-19 prior knowledge viewers have is limited, but the longer the pandemic is taking, the more knowledge people are gaining (Trajkova, et al., 2020).

I will focus on the visual aspects of graphs, knowledge respondents have about reading graphs and expectations respondents have about graphs about covid-19 cases. These factors have been studied in the context of education but not for casual data visualization. Additionally, I will look at age, gender and level of education. In this thesis, I will explore if there is also a relationship between visual characteristics, the viewer's knowledge about graphs, and the viewer's knowledge and expectations about the content in the graph and the comprehension when graphs are presented in a casual setting.

The independent variables that will be explored are visual aspects of graphs, knowledge respondents have about reading graphs and expectations respondents have about graphs about covid-19 cases and background characteristics of respondents, such as age, gender and level of education. These independent variables will be explored in their relation to the dependent variable comprehension. Comprehensions include the dimensions translation, interpretation and extrapolation. The dimensions will be used to be able to test the comprehension of the viewer. Based on the theory I expect that visual aspects of graphs, knowledge respondents have about reading graphs and level of education have an effect on the comprehension of respondents of the graphs. For the expectations respondents have about graphs about covid-19 cases I expect that viewers that are more worried about covid-19 and see the pandemic as a problem to comprehend the graphs different than viewers that are less worried about covid-19. None of the theories found discusses an effect of gender and age (within the group youth) on the comprehension of graphs and this effect is not expected to be found in regards to the research question.

4. Methodology

To explore what influences the way youth comprehends graphs about covid-19 cases, there was a need to know how these graphs are visualized. Graphs about covid-19 cases are visualized differently by different sources. For accessibility, only online sources were used in looking for graphs. The first source looked at is what the government put out itself. Secondly, different news sources, such as online news and online newspapers, were looked at. The following sources were used; NOS, RTL Nieuws, nu.nl, AD, Volkskrant, NRC, Trouw and Telegraaf. For nu.nl and Telegraaf, no graphs were found. Thirdly, when googling the amount of covid-19 cases (*aantal coronabesmettingen)* Google also gives their own graph. The graphs were selected over several days.

The relationship between the dependent variable comprehension of graphs and the several independent variables; visual characteristics, the viewer's knowledge about graphs, the viewer's expectations about the content in graph and lastly the background characteristics of the respondents, namely gender, age (in the group youth) and level of education are explored. To do this, qualitative data has been gathered through holding interviews. By gathering qualitative data respondents have more space to share their

comprehension of a graph and gives the ability to ask the respondents to elaborate on their answers. This gives a clearer view of the graph comprehension of the respondents than a survey would be able to do.

The participants were selected to have a variance in background characteristics. A variance in age, gender and education level. For the variable level of education, five different categories are focused on. MBO, HAVO, VWO, HBO and university. VMBO has been left out on purpose for finding respondents because VMBO students are often younger than 16 years old, which will limit the ability to find the right respondent. To reach people to participate in the study personal acquaintances were asked to participate. 10 respondents were found willing to participate. Before the interview took place the participant was asked about their gender, age and level of education to ensure variance in participants. The interviews were held in Dutch. Because of covid-19 regulations and easy accessibility, the interviews were held online via a video call.

During the interview, the respondents were tested on their knowledge about reading graphs by letting them answer questions. The graphs needed to answer the questions were shown on screen. After this the respondent was asked about how they see and what they think about the current situation in the covid-19 pandemic; do you think there are a lot of covid-19 cases, thoughts about regulations and if they are worried about covid-19. At last, the respondent was shown three graphs about the covid-19 cases and was asked to translate, interpret and extrapolate the graph. To do this the respondents were asked to describe the graph and tell what the graphs show, what the graph tells about the course and current situation of covid-19 in the Netherlands and how - based on the graph - they see the future of covid-19. Different graphs were made about covid-19 cases in the Netherlands, with different visual characteristics. A difference in colours, red and blue and shaded and not-shaded, and a difference in density varying in height and width. In total four red and four blue graphs were made. The respondent got to see only graphs of one colour because showing two graphs that have the same density with two different colours will look too similar for the respondent to comprehend again. Which graphs and the order they were shown were randomized before the interviews took place and were checked to see if all graphs are shown at least once. First, the colour was randomized and afterwards which three of the four graphs would be shown. In total, the respondent got shown three graphs about covid-19 cases. When showing the second and third graph, the respondent was asked to compare the graph with what they saw in the previous graph(s). Do they see differences? Does it change the way they view the previous graph(s)?

During the interview primary notes were taken, afterwards, the interview was transcribed and coded to get a better sense of the comprehension of the respondent. The data was coded by looking for content, highlighting these sentences and coding these sections. After coding, the codes were looked at to see if there were any patterns and conducting themes. For the variable knowledge about graphs, the respondent's ability to answer the question correctly will be assessed.

The results produced from the interviews cannot be generalized outside of the sample group, but gives a better understanding of the graph comprehension of the respondents and gives a starting point in developing a better understanding about how and which factors influence the graph comprehension of youth in casual data visualization about covid-19 cases.

5. Results

To answer the research question: *How do different factors influence the way youth comprehends graphs regarding covid-19 cases in the Netherlands?* two sub-questions were constructed.

- How are graphs about covid-19 cases presented in the Netherlands?

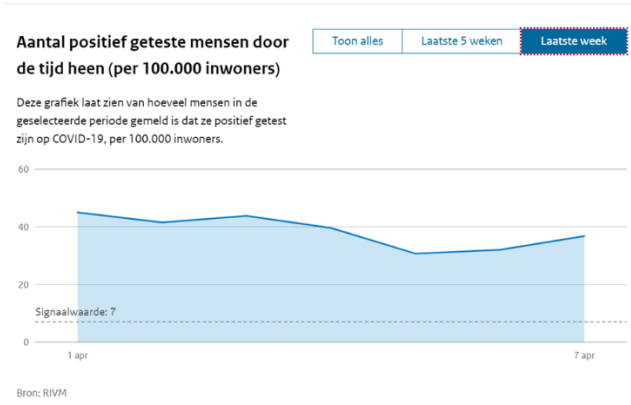
- How do visual characteristics, the viewer's knowledge about graphs, the viewer's knowledge and expectations about the content in the graph, and background characteristics of viewers influence the comprehension of graphs about covid-19?

For these sub-questions, research was completed. The results of this research will be shared below.

5.1 Ways covid-19 cases are visualized in graphs in the Netherlands

To explore what influences the way youth comprehends graphs about covid-19 cases, firstly we need to know how these graphs are visualized. Graphs about covid-19 cases are visualized differently by different sources. The different sources of the graphs are the website of RIVM, different news sources and Google. For creating their graphs. most sources either use data collected by RIVM or Johns Hopkins University. Google uses a combination of different sources, namely Wikipedia, government health ministries, The New York Times and other authoritative sources (Google, 2021)

Graphs are used to share information about quantity by using spatial characteristics. (Gillan & Lewis, 1994). The structural aspects of a graph give information about the data being measured and the kind of measurements being used. The most common type of graph is L-shaped with the y-axis representing the data being measured and the x-axis providing information about the measurements used to represent this data. In the case of graphs about covid-19 cases, the y-axis represents the number of cases and the x-axis represents the dates on which the cases were measured. The timespan of the graphs vary, some graphs only representing a small period of time, like a week (Figure 1) or a longer period of time, like several months (Figure 2).



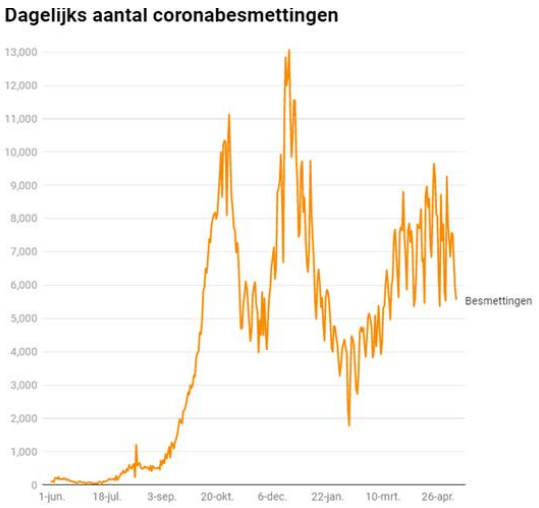


Figure 1: A graph spanning a week. Amount of positive tested people through time (per 100.000 citizens) last week. RIVM (April 7th 2021)

Figure 2: A graph spanning several months. Daily covid-19 cases Trouw (11th of May)

Visual aspects, also known as specifiers are used to represent the values of data. The specifiers can be lines in a line graph or bars in a bar graph. Graphs also have labels to indicate the type of measurements or the data to which the measurement applies or the title that tells you what the graph is about (Friel, Curcio & Bright, 2001).

5.1.1 Type of Graphs

Several different types of graphs are L-shaped and use and x- and y-axes, such as line (Figure 3), bar or a combination of bar and line graphs (Figure 4). Graphs that combines line and bars give two types of information; the number of cases and the average of the last seven days.

Figure 3: A shaded line graph. Amount of positive tests RIVM (April 7th 2021)

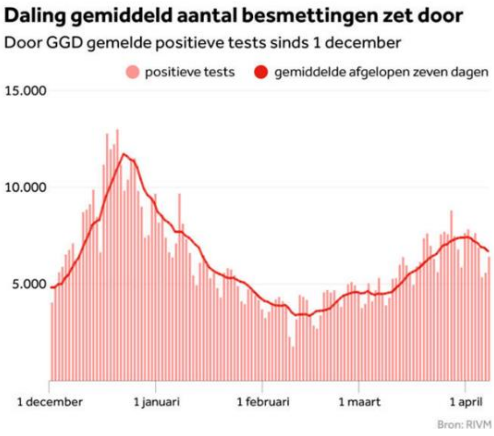


Figure 4: A graph combining line and bar. Decline average amount of infections since December 1st NOS (April 7th 2021)

Most graphs use a line, half of the time in combination with a bar. This is used when the graph is representing two variables, namely the number of cases and a seven-day average (Table 1). When only representing one variable, usually a line graph is used.

Types of graphs	
Line	5
Bar	2
Combination line and bar	5

Table 1: Frequency graph types of graphs

5.1.2 Colour

Graphs about covid-19 cases come in a variance of colours, used for lines, bars or colouring under the line. Sometimes two colours are used. (Figure 5)

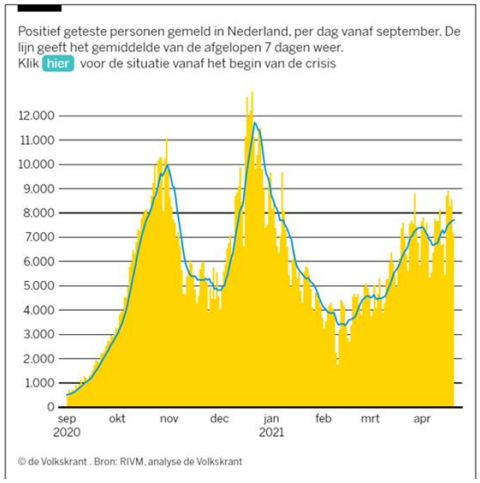


Figure 5: A graph with two different colours. The percentage of COVID-19 tests that were positive, per day Volkskrant (20th of April 2021)

Figure 6: A graph with two different colours. Positive tested people from September until now Volkskrant (20th of April 2021)

When collecting which colours are used most, sources that only used one colour were only counted once, while sources that used multiple colours in their graphs, all colours were noted. An overview of which colours were used most can be found in Table 2.

Colours of graphs	
Blue	6
Red	3
Green	1
Orange	1
Yellow	1

Table 2: Frequency table colours of graphs

The most used colour in graphs about covid-19 cases is the colour blue. After blue, red is also a colour seen in multiple graphs. The graphs that used the colour green (Figure 5) and yellow (Figure 6) was from a graph made by the Volkskrant. On the same page, several different graphs were also shown, so maybe the colours green and yellow were used to distinction the graphs from each other.

5.1.3 Density

Other than a difference in colours used in the graph, graphs from different sources also differentiate in the density of the graph. By using different sizes and aspect ratios the density of the graph is influenced, which might influence the comprehension of viewers. Figure 7 shows a graph that is tall and skinny. The line increasing and decreasing are more extreme in this graph than the graph in Figure 8, which is wider.

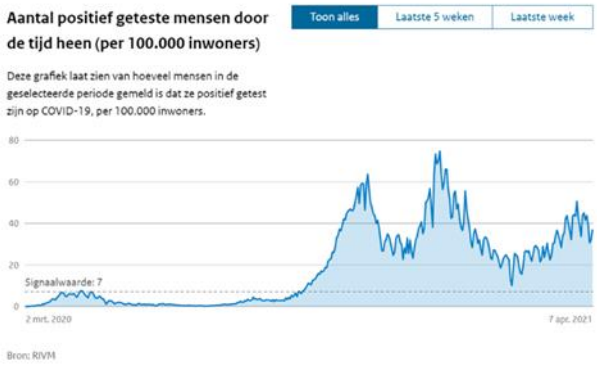


Figure 7: Positive covid-19 test and seven-day average NRC (21st of April 2021

Figure 8: A wide graph. Amount of positive tested people through time (per 100.000 citizens) all-time RIVM (April 7th 2021)

5.2 Relation between different factors and graph comprehension of youth regarding covid-19 cases in the Netherlands

After researching how graphs look like that are used to represent the number of covid-19 cases, the next sub-question can be researched: How do visual characteristics, the viewer's knowledge about graphs, the viewer's knowledge and expectations about the content in the graph, and background characteristics of viewers influence the comprehension of graphs about covid-19? The different variables will be discussed in this part.

5.2.1 Comprehension of graphs

The variable comprehension of graphs exists out of three dimensions: translation, interpretation and extrapolation. To test

R/B	References the colour red or blue
1	broad
2	Skinny and tall
3	average
4	Average and shaded

Table 3: Graph shorthand explained

5.2.1.1 translation

When translating a graph the viewer extracts information directly from the graph and translates the form of communication of the graph. In the case of the research, the respondents were asked to describe the graphs. The first graph shown was described in its complete while following up graphs

were often described by comparing the graph to the previous graph. One of the respondents, respondent 2, described the graph B2 (see Table 3 for graph shorthand explained) as "*One the y-axis we have an estimated number of contagious people. […] And then the x-axis we have different dates and that is from 2020. From February 17 2020 to April 12 2021, and we can see that, say on March 18 2020, that there were a lot of contagious people and that also went down fairly quickly between say June and August and then it went up in August again. And then it went, it was, yes, it was very high again in October and then November a little less and then December again more in January and February again less it went up again in March approximately and in April now also high again."* (Respondent 2, 170 – 181). The same respondent described the second graph they were shown (B1) as *"basically the same as the previous graph." (Respondent 2, line 234)* and describing the difference in the graphs as *´Instead of like the last graph using one date a month. This graph has two dates every month. Or not quite right. I think every two weeks is shown in this graph. It shows the same picture."* (Respondent 2, lines 234 – 237).

The first description of respondent 2 is similar to how a different respondent describes a different graph that was the first one they saw. Respondent 10 described graph R4 as: "*Well, as the headline says, the estimated number of infected. Well, I assume indeed a number of people who have had corona or I mean haven't had it, but has on such a short term that they are still contagious. Below that the dates ranging from January 3 I think to oh no, the other way around I should read April 25 – March 1. And by the looks of it, it starts at 20.000 and ends at 200.000, I mean. And, we're very much seeing that peak in March. Indeed, what did they say, a kind of a third wave followed by a very large decline, which is immediately noticeable. And after that, it actually continues to fluctuate a bit, but still quite high so yes, it strikes me right away that despite the decrease in yes, I am now in all of course already in June. If you're looking for an April 20, we already at 180. That's even more than in March."* (Respondent 10, lines 122 – 130). The other eight respondents described the first graph they saw similarly as respondents 2 and 10, but used extrapolation in their description.

5.2.1.2 interpretation

Interpretation asks viewers to rearrange material and sorting the important data from the less important data. To do this viewers can look for a relationship in a graph. For interpretation, the viewer needs to use two or more pieces of data. In this, respondents were asked how they found the current covid-19 situation was going. For this question, the respondents had to compare several points in the graph to the last point in the graph. Respondent 4 said: *'Yes, you can see that infections are really jumping up now."* (Respondent 4, line 154) in regards to the question, comparing the last point in the graph to earlier points in the graph. Another respondent answered with: *"Yes it does go down, but if you look at the foregoing, it was lower all the time, It certainly isn't going very well."* (Respondent 6, lines 144 – 145).

Respondents also used interpretation for comparing how several points follow each other up in the graph. When comparing graphs R3 and R2, respondent 10 described graph R2 as *'It's still going up really steep. Actually, it also seems much steeper, it fluctuates more. Then you're like okay, where – where is that coming from, because of a drop. It goes up really quickly again, so not such a good impression of it either."* (Respondent 10, lines 179 – 182). When comparing points in graph R2 the respondent saw a steeper follow up than in graph R3.

Another respondent compared the graphs R1 and R2 and said: *"that previous chart [R1] that looked like, like what that high and low were closer together. And well, that previous chart that looked like, but what that high and low were closer together. And here [R2] it is, or it's really all the way up. Or It's really all the way down. You can see clear differences."* (Respondent 5, lines 205 – 207). And when asked whether that changed the way they saw graph R2, respondent 5 said: *´ Oh yes well, here it*

seems like something, worse or something like that […]. Kind of the extreme version of the previous one, I think." (Respondent 5, lines 213 & 217). In this case, the respondent did not compare the steepness of the line in the two graphs but the space between two points.

5.2.1.3 extrapolation

For extrapolation, the viewer needs to use information beyond the graph they were shown. As said before, in the case of the graphs about covid-19 cases the respondents used a lot of extrapolation in their description of graphs. Respondent 1 described graph R3 as: *"I see it at the beginning of 2020, the corona really first appeared in the Netherlands. At that time, the Netherlands was not doing a lot of measures. And as the measures were kept up, the people went into lockdown. The contamination has gone down very much. But in the end, they thought, let's open up a little bit again, which was a small mistake. And that ultimately resulted in many infections and you can clearly see that there are several waves. And we're going down again."* (Respondent 1, lines 159 – 164).

Another respondent described graph R1: "*This is early 2020. That's when it came in I think. Then came a very big [wave] yes. Then it was probably a lot of people infected because there were no measures yet. When were the measures taken? And then they went very well for a very long time, but this is in the summer. So I think that's weird. But okay in the summer there were apparently no infections or well, very few in any case when we look, then it was 11, 12 and then in the winter it increases a bit and but that is quite weird because logically, the measures were also increased again, so I think that's weird. But maybe it could also be that people thought oh, we don't care anymore, so we don't stick to it. And then it became more [cases]."* (Respondent 5, lines 145 – 156). Even though respondent 5's background information did not match the graph that they were shown, they still used the background information to describe the graph, filling in the holes with possible explanations why the background information they had did not match. The respondents that used extrapolation in their description of the graphs did so to explain why the graph increased and decreased.

To answer the question of how they thought the current covid-19 situation was going, conducted to see how the respondents would interpret, respondents also used extrapolation in their answers. For example respondent 2 answer the question with: *"I think the government is doing their very best to open everything up again, because people will whine otherwise, but that, says that opening everything does ensure that there are more infections again. Even though in this case it is less dangerous and most infections are no longer very disease-carrying. Or, sick? […]. [People are less sick of it than last year?] Yes, that indeed. Of course, also something that in March or April [we] started vaccinating. I don't know how much that would affect this graph. But that's something to keep in mind as always."* (Respondent 2, lines 204 – 214.)

Respondents were also asked about what they expected would happen to the amount of covid-19 cases. Two types of answers were found with the respondents. Five respondents expected the number of cases to go down and rise again, just like last year's summer. Respondent 4 said: *´I think that we may soon have a very big decrease again, that then everything, just like last time will be opened up again. And that we will get an exceptional increase again."* (Respondent 4, lines 165 – 167). Four respondents expected a decrease because: *´we're busy now or that there are now quite of a lot of vaccinated, eh, I think it will go down. I don't think it will go down very quickly. But yes, that it does get – yes, it will decrease."* (Respondent 6, lines 152 – 154).

While all three dimensions of comprehension are present in the comprehension of the respondents interviewed, extrapolation is used a lot, even though to answer some of the questions only using translation or interpretation would have been enough.

5.2.2 Visual characteristics of graphs

Using two different colours, red and blue, four different types of graphs were made. The four graphs varied in density, one being really tall, one being really wide and two being average. One of the average graphs was shaded, instead of only using a line (Appendix B.2). The ten respondents were asked to describe three graphs and compare the graph to the one shown before and if the new graph changed the way they saw the previous graph(s). Five of the respondents got red graphs, the other five respondents got blue graphs.

Interview	First graph	Second graph	Third graph
1	R3	R1	R4
2	B2	B1	B3
3	B1	B2	B4
4	R4	R3	R1
5	R1	R2	R4
6	B2	B4	B1
7	R3	R1	R2
8	B3	B2	B1
9	B1	B2	B4
10	R4	R3	R2

Table 4: Which graph got shown to which respondent

5.2.2.1 Colour

The respondents that were shown red graphs talked more about how some of the graphs looked aggressive or scary compared to the other graphs they were shown than the respondents that were shown blue graphs. Respondents that were shown red graphs described differences between graphs as the following; *"This graph [R2] - Oh, yes, well, here it seems like something, a bit worse or something like that here I yes. [...] Bit the extreme version of the previous one [R1], I think."* (Respondent 5, lines 210 – 215) Respondent 7 compared graphs R3 and R1, with R1 being a wider graph. *"Yes the same, actually a bit like the previous chart [R3], but then with a look, I think much more different, how do you say that. Say much difference more information say. It [R1] is much more specific in that sense, say you have. I don't know how to say this."* (Respondent 7, lines 147 – 149). For respondent 10, the steepness of the graph was noticed when comparing graph R2 to the two earlier graphs they were shown R4 and R3, which changed their impression of the graph. *"Yes, exactly. Still seen here. You think, oh, it's still going up really steep. Actually, it also seems much steeper, it fluctuates more. Then you're like okay, where - where is that because of a drop but. It does go up really fast, so not a good impression of that either."* (Respondent 10, lines 179 – 182).

Respondents who got shown a blue graph described graphs as aggressive way less and were more likely to just describe a translation difference. For example, when asked to describe B1, they said: *"This is showing the same graph as the previous two [B3 and B2], only now very wide."* (Respondent 8, line 185). And when asked if this graph changed the way they saw the previous graphs, they answered *"no."* (Respondent 8, line 188). This does not mean that none of the respondents who were shown a blue graph described a graph as aggressive when comparing. Respondent 9 said when comparing B1 and B2: *"No yes, maybe that first peak. It comes out even more clearly now, making it more frightening."* (Respondent 9, lines 136 – 137).

Two of the graphs that were shown to respondents were shaded in, one red and one blue. Graphs that were shaded, when asked if that changed the way they saw the previous graphs(s), were described as the following. "*Yes, a bit, because now it seems as if that, the high number of infections at the last peak, will actually remain for a while because [a] large wall - blue wall almost comes into view."*

(Respondent 9, lines 149-150). And *"This one [R4] seems scarier. On one side because it is all red instead of a small line."* (Respondent 1, lines 273).

Respondents that were shown the shaded graphs before other graphs gave these comparisons: *"Well in the first instance in terms of dates, it seems very much the same, [but] if you really pay attention to details, such as the start and end date, then there are differences. But yeah, I don't really do much with that. And still a lot the number is still very high and if you think about it indeed. Still too high and would probably also lead to more infections. But it's since that is just a line and not that very big filled area. It seems less threatening or less telling."* (Respondent 10, lines 157 – 162). And *"Yes, these graphs [R3 & R1] look a little less intense so to speak than the first [R4]."* (Respondent 4, lines 210).

Red graphs seem to make respondents describe graphs as more aggressive or extreme when comparing graphs with different densities. This happened less when the graphs were coloured blue. Graphs that were shaded were described as scary and threatening. The shaded graph also gave the idea that the number of cases will remain for a longer time, than a graph that just has a line.

5.2.2.2 Density

There were three different graphs created with a variance in density. The wide graph was described as more clear and specific.

The widest graph gets described by respondents as less aggressive, less intense and clearer. Respondent 6 described graph B1, the widest graph in comparison to graphs B2 and B4 as: *"Yes yes, because this yes is more. It's a bit wider than the previous one. And yes it looks less aggressive [than] those peaks. They look yes, which yes, should be less aggressive, so to speak. Yes, it's a little more gradual."* (Respondent 6, lines 171 – 174). Respondent 4 says: *"Yes, these graphs [R3 & R1] look a little less intense so to speak than the first [R4]."* (Respondent 4, line 210). Respondent 7 noted the wide graph as: *"Yes the same, actually a bit like the previous chart [R3], but then with a, I think much more different, how do you say that. Say much difference more information say. It is much more specific in that sense, say you have. I don't know how to say this. [...] It's clearer, you can see more or something."* After seeing the wide graph first and afterwards a skinnier graph respondent 7 said: *"Yes yes I think so. I think I'm going to look at that graph differently now. Indeed that I say but in the back of my mind is still okay, but there is also more information that I still miss and that I still look somewhat suspiciously at that other graph."* (Respondent 7, lines 167-169).

The tall and skinny graph was described in comparison to other graphs as leaving a bad impression, more frightening and showing a clearer difference in the points in the graph. Respondent 10 noticed that the tall graph was steeper: *"You think, oh, it's still going up really steep. Actually, it also seems much steeper, it fluctuates more. Then you're like okay, where - where is that because of a drop but. It does go up really fast, so not a good impression of that either."* (Respondent 10, lines 179 – 182). Respondent 5 said that the difference looked bigger: "*And here (R2) it is, or it's really all the way up. Or It's really all the way down. You can see clear differences."* (Respondent 5, lines 206 – 207). Respondent 9 also noticed the bigger difference in the low and high points and said: *"No yes, maybe that first peak. It comes out even more clearly now, making it more frightening."* (Respondent 9, lines 136 – 137).

Changing the aspect ratios, which in turn changed the density of the graph, changes the comprehension of the graph. Mostly the interpretation is influenced. A skinny and tall graph is seen as scarier than a wider graph, while the wide graph is seen as more specific and clear.

5.2.3 Viewer's knowledge about graphs

To test the knowledge of respondents, 9 questions about graphs were asked (Appendix A). The number of questions answered correctly is used to assess the knowledge of the respondent.

Only one respondent scored low on knowledge about graphs, with a 4,5 out of 9 (Respondent 1). Three respondents scored 7 or 7,5 and six respondents scored 8 or 9 (Table 5), making it hard to analyse how a viewer's knowledge about graphs influences their comprehension. Nevertheless, the answers between the respondent who scored the lowest and the two respondents who scored the highest were compared.

Score on the knowledge test	
4,5	1
7	2
7,5	1
8	4
9	2

Table 5: Frequency table score on the knowledge test

When comprehending the graph respondent 1 (who scored low) used extrapolation a lot when describing the graphs. Graph R3 was described as: *"I see it at the beginning of 2020, the corona really first appeared in the Netherlands. At that time, the Netherlands was not doing a lot of measures. And as the measures were kept up, the people went into lockdown. The contamination has gone down very much. But in the end, they thought, let's open up a little bit again, which was a small mistake. And that ultimately resulted in many infections and you can clearly see that there are several waves. And we're going down again."* (Respondent 1, lines 159 – 164) and graph R1 was described as: *"At the beginning when the corona broke out, we did a very good job to contain the corona. But as the relaxation came again, people took corona a lot less seriously, because a lot of people have had it. And in the end, they didn't give a dam, which means that the infections are now slightly higher than before."* (Respondent 1, lines 226 – 229).

Respondents 6 and 10, who scored 9/9 on the knowledge test described the graphs with more usage of translating. Respondent 6 described graph B2 as: *"Well there are a lot of peaks, but also quite a lot of throughs in between, or yes, it started in February 2020 and then the first peak was in March, but it fell very quickly. Probably a month later in April Because then the first measures were taken, the lockdown was quite strict then. Well, then they stayed low for a while and then in August, late August, early September, I think. The contamination went up again and there comes the second peak, which is slightly less high than the first peak. Yes and then it went down again In the coming months and in December it also went up a bit. With very, very slightly above the peak, the second peak of September. Yes. And then it descended again and now yes now it drops at the end it drops again a little bit, but it has already risen quite a bit."* (Respondent 6, lines 125 – 133). While there is some extrapolation in this answer, namely mentioning the measures and lockdown, this is a way less extent than respondent 1 had done. Another graph respondent 6 described, graph B4 contained no extrapolation. *"Well in, March 2020 it went up really fast and then it was at the peak at, ok in March, yes, also in March. That's the first peak and then it fell back quickly. But then it stayed low until end of August 2020. Then came another peak then, then came another peak. This one is yes even a bit like the chart before but a little longer. So then I went down a little bit and then up again and then down and then up again and down a little bit."* (Respondent 6, lines 153 – 157). Respondent 10 also did not use extrapolation in their description: *"Well, as the headline says, the estimated number of infectious. Well, I assume indeed a number of people who have had corona or I mean haven't had it, but has on such a short term that they are still contagious. Below that the dates ranging from January 3 I think to oh no, the other way around I should read April 25 – March 1. And by the looks of it, it starts at 20.000 and ends*

at 200.000, I mean. And, we're very much seeing that peak in March. Indeed, what did they say, a kind of a third wave followed by a very large decline, which is immediately noticeable. And after that, it actually continues to fluctuate a bit, but still quite high so yes, it strikes me right away that despite the decrease in yes, I am now in all of course already in June. If you're looking for an April 20, we already at 180. That's even more than in March." (Respondent 10, lines 122 – 130)

A low knowledge of graphs seems to mean that the respondent uses more extrapolation in their description of a graph. Respondents that have a high knowledge of graphs mainly used translations in their descriptions. This finding was surprising because extrapolation is seen as more difficult to do than translating.

5.2.4 Viewer's expectations about the content in the graph

Respondents were asked to tell about the covid-19 situation in the Netherlands and how worried they were about covid-19.

Worrisome	
Not worried	3
Not worried for themselves	3
Worried	4

Table 6: worrisome of respondents about covid-19

Respondent 3 describes the covid-19 situation as the following: *"Well, you've had the first wave, you've had the second wave, you've had and now we've got the third like that. And well, if I may be quite honest, me. Think that. Umm, I think the government is quite underestimating the corona infections. Now that they're opening up all the stuff, because I know people who work in the hospital and the ICU, they're all still full. And now they're just opening things up again. Yeah, I don't think that goes well."* (Respondent 3, lines 98 – 102). And when asked if they were worried about covid-19 they said: *"Well I do a little bit. That's also because I have a small medical background. And, I have no idea what corona would do to me."* (Respondent 3, line 133). Respondent 3 described graph B1 as the following: *"Its first was the zero infections of humans. Then suddenly it went around March? Yes, that's March 2020. They suddenly went very fast, they were about one hundred and sixty thousand. Infections then suddenly it went down very slowly, very quickly low. And then they suddenly rose again in August? Yes, that's August. And then it kept fluctuating a bit again. Rise and fall and rise and fall and now it rises again and we are in January 2021, er April 2021."* (Respondent 3, lines 152 – 156). When asked if they thought the current situation was good or bad they said: *"Yes, well, me, I'm glad the numbers are falling, but I'm afraid of the consequences that will come after all the terraces are open and you can only see that in two weeks."* (Respondent 3, lines 179 – 180). The respondent also thought the number of cases would rise again. (Respondent 3, lines 188).

Respondent 8 was not worried about covid saying: *"Because I. Yeah personally it doesn't bother me that much. I have not been exposed to [covid]. I know about my parents, that's how I've had it, my parents and my sister, But I've actually cycled through it completely and I actually didn't suffer from it at all. As far as I know, I have never been infected. I have the feeling that if that was the case, […] I don't know exactly, but I have a feeling it should be fine."* (Respondent 8, lines 111 – 115). When asked if there were a lot of cases they answered: *"By the way, I didn't delve into that at all I think. Logically I would say, We are going to relax, so there are fewer infections, but I think it is still disappointing how many infections there are."* (Respondent 8, lines 79 – 81). When asked to describe graph B3 the respondent said the following: *"Well at the bottom you have a time period, then you see that, what is? The first corona cases started to show up in mid-February. Then you were really, March April that first peak so that all flattened out a bit. And then it started again around November. Yes,*

October - November has really increased considerably. Little bit decreased later we increased it, bit decreased again. And now we are actually increasing." (Respondent 8, lines 122 – 126). When asked if the current situation was good or bad they said: *"No, no, I had expected that we would still be high. I actually expected that third peak to say a little lower, because it was sixty thousand or so and that we now once because I knew we had another peak that we were now at our hundred thousand or something. But it's almost double it."* (Respondent 8, lines 134 – 137).

Respondent 8 was not the only respondent who was surprised by the number of current cases. Respondent 7 said: *"And yes if I would see the numbers now. uh then yes. That is somewhat worrying."* (Respondent 7, line 133) Respondent 10 was also surprised by the number of cases. *"Well, if I had to base myself purely on this chart, I'd say, this isn't going in the right direction. If I say two thousand infections is good, but I see that there are just one hundred and eighty thousand contagious. Then I start to worry a bit."* (Respondent 10, lines 134 -136). Both respondent 7 and respondent 10 were somewhat worried about the cases being higher than they expected.

Another respondent whose expectations did not meet the graph was respondent 5. *"Likewise looking, This is early 2020. Then it came in I think. Then came a very big yes. Then it was probably a lot of people infected because there were no measures yet. Then the measures were taken and then they went very well for a very long time, but this is in the summer. So I think that's different. But okay in the summer there were apparently no infections or well, very few in any case when we look, then it was 11, 12 and then in the winter it increases a bit and But that is quite different, because if all goes well, the measures were also increased again, so I think that is strange."* (Respondent 5, lines 138 – 144). Before seeing a graph respondent 5 described the current amount of cases as *"Um well, I could imagine since it's summer and [people are] getting together now. Yes, so I think I do think there are quite some [cases]."* (Respondent 5, lines 97 – 98). After seeing the graph that showed that the cases in the summer of 2020 were actually really low, the respondent expected that the number of cases would lower, saying that *"Um yeah, if we're just going to have this whole process like that if we're going to continue that, just say it subsides over the summer. Just like last year, then I think it will go down again, but of course, that is a totally different summer and We already know a lot more about it. Well, yes I don't, but then of course, so the virus much more researched now and the vaccinations. I also hope it goes down again. Yes, I think so too then."* (Respondent 5, lines 173 – 177).

When comparing the answers given by respondents who said they were worried before seeing the graphs with the respondents who replied they were not worried, no difference was found in the comprehension. Not in translating, interpreting or extrapolating. When graphs did not look like the respondent expected the graph to look like, respondents were surprised but were able to change their views, by changing the extrapolation in their description. Two respondents who expected a lower number of cases were somewhat worried after.

5.2.5 Background characteristics of viewers

Three background characteristics were asked of the respondents beforehand. Gender, age and level of education (Table 7: Frequency table background characteristics).

Level of education	
MBO	2
Havo	2
VWO	2
HBO	3
University	1
Gender	
Male	4

Female	6
Age	
17	3
18	2
19	4
20	0
21	1

Table 7: Frequency table background characteristics

The answers respondents gave about graph comprehension was compared with the background information of the respondents to see if there were possible relations between the variables.

The three different variables were compared with how often the three different dimensions of comprehension – translation, interpretation and extrapolation – appeared in the descriptions respondents gave about graphs. Whether the different types of graphs made them change the way they saw previous graphs and whether they thought the current covid-19 situation was good or bad was not dependent on the three background characteristics. There were no real differences found between gender, age and level of education.

6. Conclusion and discussion

This research was conducted to answer the question: *How do different factors influence the way youth comprehends graphs regarding covid-19 cases in the Netherlands?* To do so, several variables have been studied in their relation to comprehension of graphs about covid-19 with the three dimensions; translation, interpretation and extrapolation. The independent variables studied are visual characteristics of graphs, viewer's knowledge about graphs, viewer's expectations about the content in the graph and the background characteristics of viewers; gender, age and level of education.

To answer the research question *How do different factors influence the way youth comprehends graphs regarding covid-19 cases in the Netherlands?*, three variables were found as factors that influence the comprehension of youth about graphs. These were the visual characteristics of graphs, the expectations about the graph and the knowledge the viewer holds about graphs. Especially the visual characteristics of the graphs seem to have a strong influence. If you want to make a graph that is scary; a red, shaded, skinny and tall graph will do so. On the opposite end would be a blue and wide graph, this graph would be considered clear and specific. But most of all, viewers of the graph seem to rely a lot on background information they hold about the covid-19 pandemic, using extrapolation in their description of the graphs.

Before conducting the research I expected that visual aspects of graphs, knowledge respondents have about reading graphs and level of education would have an effect on the comprehension of respondents of the graphs. For the expectations respondents have about graphs about covid-19 cases I expected that viewers that are more worried about covid-19 and see the pandemic as a problem to comprehend the graphs different than viewer that is less worried about covid-19. I did not expect gender and age (within the group youth) to have an effect on the comprehension of graphs.

During the comprehending of the graphs about covid-19, the respondents used extrapolation a lot during their description of the graphs. The respondents used the background information they had about the covid-19 pandemic to explain why a graph was rising or falling. When guidelines were not strict enough you could see a rise and when new guidelines were just implemented the graph was decreasing.

The visual characteristics of the graphs seem to influence the comprehension of the graphs about covid-19. The respondents that were shown red graphs talked more about how some of the graphs

looked aggressive or scary compared to the other graphs they were shown than the respondents that were shown blue graphs. The graphs that were shaded were described as scarier and more aggressive than the graphs that were not shaded. The wide graph was described as more clear and specific. One respondent said that the broad graph makes the low cases in summer more noticeable than in a skinnier graph. The tall graph is described as more aggressive and scary. The average graph is described as more aggressive than the wide graph, but less worrying than the tall graph. Concluded, the wider the graph, the clearer. Furthermore, the skinnier and taller the graph, the scarier and more aggressive. That the visual characteristics of a graph would influence the comprehension was expected.

Between the respondent that scored the lowest and the two respondents that scored the highest for their knowledge test, there seems to be a difference in the way they described their graphs. The respondent that had a low score on the knowledge test used a lot of extrapolation in their description of the graphs, while the two respondents who scored high on the knowledge test used a lot more translation and less extrapolation. As expected the knowledge the viewer had about graphs influences the comprehension, but this finding was still surprising because theory stated that extrapolation is the most challenging, but the findings of this research about graphs about covid-19 show the opposite. The respondents that scored high on the knowledge test did not show difficulty with extrapolation when answering questions, but the respondent who scored low on knowledge seemed to need to rely on background information to describe the graph. Another explanation for this might be that the respondents that had a high knowledge interpreted the question *can you describe this graph?* differently than the respondent with low knowledge about graphs.

The worry respondents had about the current covid-19 situation did not seem to influence the comprehension of graphs when describing or comparing graphs or the way they expected the number of cases to continue. If the expectations respondents had did not match with the graphs, they were surprised but were able to change their views. By changing or adding to the extrapolation in their comprehension, they were able to make the background information fit. Two respondents who expected a lower number of cases were somewhat worried after. It should be questioned whether the respondents who said they were not worried, were not worried about the number of cases or that they themselves would get covid and get sick. This finding was not what was expected before conducting the research.

For the three different background characteristics studied, no difference was found in any of the dimensions of comprehension or description in the differences between graphs. For age and gender, no influence was expected and this is also the case. For level of education, an influence on the graph comprehension was expected but not found.

For this research only a small sample group was used, meaning a limit in the generalizability of the research findings. Further research is needed to establish the possibility for generalization. Especially the variable knowledge about graphs was hard to study in the relation of the comprehension of graphs about covid-19. A bigger sample group will result in a larger variance in a viewer's knowledge about the graph and other variables. When interviewing respondents it happened that some respondents immediately recognized that the graphs were showing the same data just with different aspect ratios. For some of these respondents, this meant that they were not able to comment or did not see a difference between the graphs.

7. Literature

Binnenlandredactie AD. (2021, April 15). *TERUGLEZEN | 538 Oranjedag stijf uitverkocht, fors meer nieuwe coronagevallen door technische storing*. AD. https://www.ad.nl/binnenland/teruglezen-538-oranjedag-stijf-uitverkocht-fors-meer-nieuwe-coronagevallen-door-technische-storing~ac90fcc4/

Brockmann, R. J. (1991) "The unbearable distraction of color," in IEEE Transactions on Professional Communication, vol. 34, no. 3, pp. 153-159, Sept. 1991, doi: 10.1109/47.84109.

Friel, S., Curcio, F., & Bright, G. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. Journal for Research in Mathematics Education, 32(2), 124-158. doi:10.2307/749671

Gal, I. (1993). Reaching out: Some issues and dilemmas in expanding statistics education. In Pereira-Mendoza (Ed.), Introducing data-analysis in the schools: Who should teach it and how? 189-203). Voorburg, The Netherlands: International Statistics

Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. Human Factors, 36, 419-440.

Google. (2021, April 9). *Coronagevallen - Google zoeken*. https://www.google.com/search?sxsrf=ALeKk02jNtGE_K1YBewesXPm3zi-QGp2-w:1617961599851&q=coronagevallen&spell=1&sa=X&ved=2ahUKEwic5qzT8PDvAhXf_7sIHVsUDW0QBSgAegQIARA2&biw=1280&bih=578

Google. (2021) *Coronavirus (COVID-19) statistics data – about this data* https://support.google.com/websearch/answer/9814707?p=cvd19_statistics&hl=en-NL&visit_id=637564295109562307-3022028063&rd=1

Henn, M. & Foard, N. (2014). Social differentiation in young people's political participation: the impact of social and educational factors on youth political engagement in Britain, Journal of Youth Studies, 17:3, 360-380, DOI: 10.1080/13676261.2013.830704

Johns Hopkins University & Medicine (2021, May 12) *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)* Coronavirus Resource Center https://coronavirus.jhu.edu/map.html

Ministerie van Algemene Zaken. (2020, 7 augustus). Letterlijke tekst persconferentie minister-president Rutte en minister De Jonge (6–8-2020). Mediatekst | Rijksoverheid.nl. https://www.rijksoverheid.nl/documenten/mediateksten/2020/08/06/letterlijke-tekst-persconferentie-minister-president-rutte-en-minister-de-jonge-6-8-2020

Mathieson, S. A. (2021, March 2). Covid-19 and the art and science of data visualization. ComputerWeekly.com https://www.computerweekly.com/feature/Covid-19-and-the-art-and-science-of-data-visualisation

NOS. (2020, March 29). *Corona in Nederland: cijfers van 29 maart*. https://nos.nl/artikel/2327537-corona-in-nederland-cijfers-van-29-maart

NOS. (2021, April 7). *6409 nieuwe besmettingen, daling zet vooralsnog door*. https://nos.nl/artikel/2375763-6409-nieuwe-besmettingen-daling-zet-vooralsnog-door

Rijksoverheid. (2020, 6 augustus). 6 augustus 2020: Integrale persconferentie van premier Rutte en minister de Jonge [Video]. YouTube. https://www.youtube.com/watch?v=wHFxmdr5VPw

Rijksoverheid. (2021, April 8). *Coronadashboard | COVID-19 | Rijksoverheid.nl*. Coronadashboard. https://coronadashboard.rijksoverheid.nl/

Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. Health Economics, 29(11), 1482–1494. https://doi.org/10.1002/hec.4143

RTL Nieuws. (2021, April 6). *Aantal coronabesmettingen daalt licht deze week, ic-opnames blijven stijgen*. https://www.rtlnieuws.nl/nieuws/nederland/artikel/5223799/rivm-aantal-coronabesmettingen-stabiliseert-deze-week-ic-opnames

Shah, P., & Freedman, E.G. (2009). Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. *Topics in Cognitive Science, 3*(3), 560-578. https://doi.org/10.1111/j.1756-8765-209.01066.x

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. Educational Psychology Review, 14(1), 47–69. https://doi.org/10.1023/A:1013180410169

Sprague, D., & Tory, M. (2012). Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. Information Visualization, 11(2), 106–123. https://doi.org/10.1177/1473871611433710

Szafir, D. A. (2018). Modeling Color Difference for Visualization Design. IEEE Transactions on Visualization and Computer Graphics, 24(1), 392–401. https://doi.org/10.1109/tvcg.2017.2744359

Trajkova, M., Alhakamy, A. ’., Cafaro, F., Vedak, S., Mallappa, R., & Kankara, S. R. (2020). Exploring Casual COVID-19 Data Visualizations on Twitter: Topics and Challenges. Informatics, 7(3), 35. https://doi.org/10.3390/informatics7030035

Trouw. (2021, May 11). De belangrijkste coronavirusstatistieken op een rij. https://www.trouw.nl/buitenland/de-belangrijkste-coronavirusstatistieken-op-een-rij~bcd4d8cd/

Tunçgenç, B., El Zein, M., Sulik, J., Newson, M., Zhao, Y., Dezecache, G., & Deroy, O. (2021). Social influence matters: We follow pandemic guidelines most when our close circle does. British Journal of Psychology. Published. https://doi.org/10.1111/bjop.12491

van Gameren, E. (2021, April 21). *Zo is het coronavirus nu verspreid over Nederland*. NRC. https://www.nrc.nl/nieuws/2020/09/29/zo-is-het-coronavirus-nu-verspreid-over-nederland-a4014030

van Lochem, T., & van Mersbergen, S. (2021, May 8). Coronacijfers op de kaart en in grafieken: waar zijn veel besmettingen? AD. https://www.ad.nl/binnenland/coronacijfers-op-de-kaart-en-in-grafieken-waar-zijn-veel-besmettingen~a4545530/

van Uffelen, X., Frijters, S., Ajrović, S., & Hofstede, S. (2021, April 21). *De belangrijkste grafieken en kaarten over de uitbraak van het coronavirus in Nederland*. de Volkskrant. https://www.volkskrant.nl/nieuws-achtergrond/de-belangrijkste-grafieken-en-kaarten-over-de-uitbraak-van-het-coronavirus-in-nederland~b18f4613/

Zitek, E. M., & Schlund, R. J. (2021). Psychological entitlement predicts noncompliance with the health guidelines of the COVID-19 pandemic. Personality and Individual Differences, 171. https://doi.org/10.1016/j.paid.2020.110491

8. Appendices

Appendix A: Interview Questions

Hallo,

Ik ben een student aan de Universiteit van Twente en doe onderzoek naar hoe het aantal coronabesmettingen wordt weergegeven en hoe jongeren daar naar kijken. Voor dit onderzoek zou ik een aantal dingen aan jou willen vragen over jou zelf en over grafieken, hoe lees je de grafiek en welke grafieken hebben jouw voorkeur. De gegevens die ik verzamel zullen anoniem zijn, veilig bewaard worden en alleen gebruikt worden voor dit onderzoek. Als je je wilt terugtrekken van het onderzoek is dit mogelijk.

Voordat we beginnen aan dit interview wil ik jou vragen of je toestemming geeft voor de volgende dingen:

- Het verzamelen en verwerken van jouw gegevens die je invult in de enquête
- Het opslaan van deze gegevens
- Het publiceren van de gegevens in mijn onderzoek
- Dit interview opnemen

Als je contact met mij wil opnemen over het onderzoek kan dat.

Contact:

Anke Morsing – Student Management, Society and Technology

a.morsing@student.utwente.nl

Testing knowledge:

Vragen bij figuur 1

1. Hoeveel kilometer heeft de intercity afgelegd na 10 minuten?
2. Op hoeveel stations stopt de stoptrein?
3. Hoeveel kilometer heeft de intercity na 30 minuten meer afgelegd dan de stoptrein?
4. Hoe lang staat de intercity in totaal stil?

Vragen bij figuur 2:

5. wat wordt hier afgebeeld?

Vragen bij figuur 3:

6. wat wordt hier afgebeeld?
7. Hoeveel meer honden dan katten zijn er?

Vragen bij figuur 5:

8. hoeveel fietsen waren er in Nederland in 1980?
9. Hoeveel fietsen verwacht je in 2030 gebaseerd op deze grafiek?

Expectations:

Kun je mij wat vertellen over hoe de corona nu verloopt?

- Vind je dat er nu veel besmettingen zijn of niet?
- Wat vind je van de huidige maatregelen? (Zijn ze te streng of te soepel?)
- Vind je dat de maatregelen helpen met het verminderen van het aantal corona patiënten.
- Denk je dat mensen zich goed aan de maatregelen houden?
- Maak je je zorgen over corona?

Visual aspects:

From the graphs on the next page, three different ones will be chosen to be shown to respondents.

Translation (*Translating a graph requires the viewer to change the form of communication*):

- Kun je de grafiek omschrijven?
- Wat wordt er in de grafiek weergegeven?

Interpretation (*viewers to rearranging material and sorting the important data from the less important data*):

Wat zegt de grafiek over het verloop van corona in Nederland?

- Zijn er periodes waar het beter/slechter gaat
- Gaat het in het algemeen goed nu?

Extrapolation (*extends on the interpretation of graphs and not only looks at what the graph is depicting but also possible consequences by looking for trends or implications*):

- Gebaseerd op de grafiek, wat denk je dat er met het aantal coronabesmettingen in de komende tijd gaat gebeuren?

When showing the next graphs:

- Wat kun je deze grafiek omschrijven?
- Zie je verschillen tussen de twee grafieken/de vorige grafieken?
- Verandert deze grafiek, hoe je de eerste/eerdere grafiek(en) ziet?

Answers for testing knowledge:

1. 9 km
2. 4 stations
3. 9 km
4. 5 minuten
5. percentages
6. absoluut/totalen
7. 75
8. ongeveer 11 miljoen
9. ongeveer 26 miljoen

Appendix B: Figures used in the interview

1. Graphs for testing knowledge

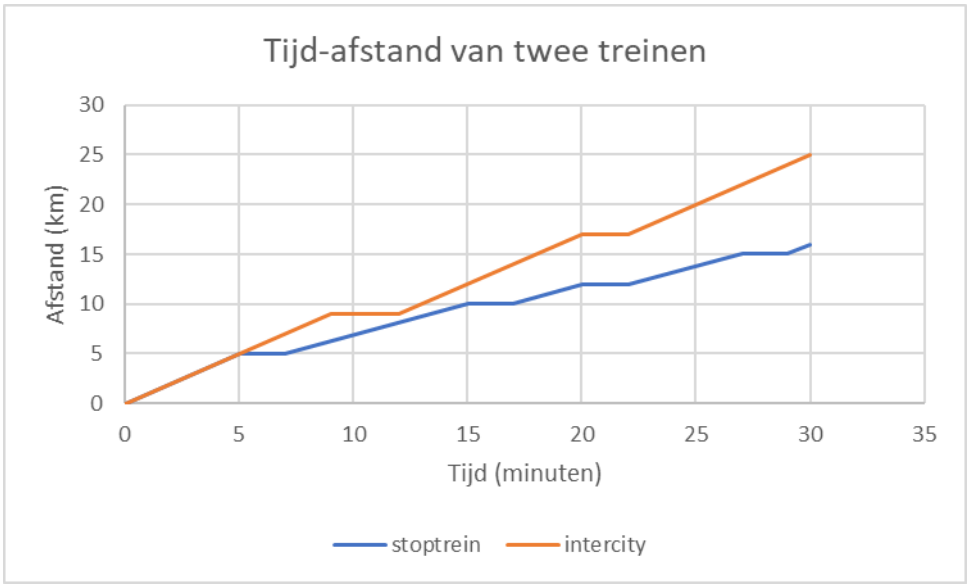


Figure 9: Tijd-afstand van twee treinen

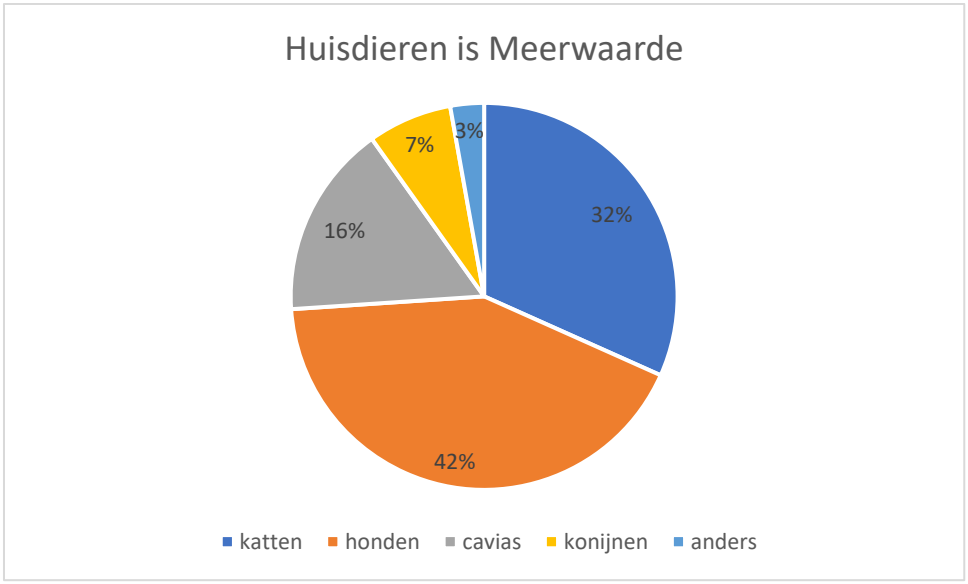


Figure 10: Cirkeldiagram huisdieren in Meerwaarde

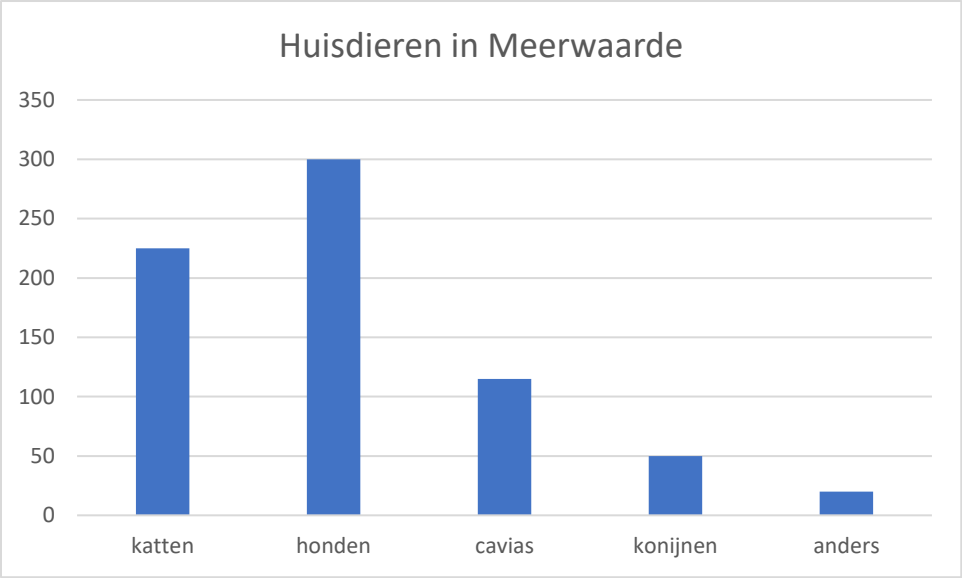


Figure 11: Staafdiagram huisdieren in Meerwaarde

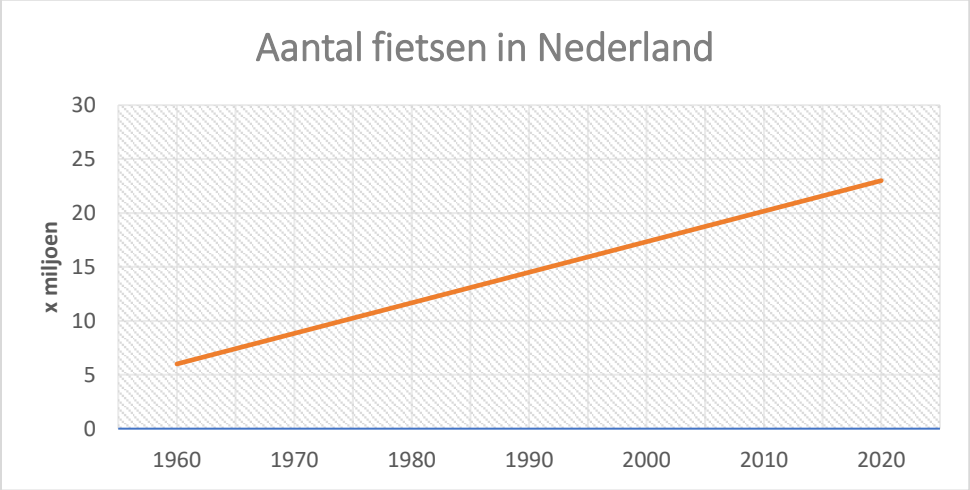


Figure 12: aantal fietsen in Nederland

2. Graphs about covid-19 cases

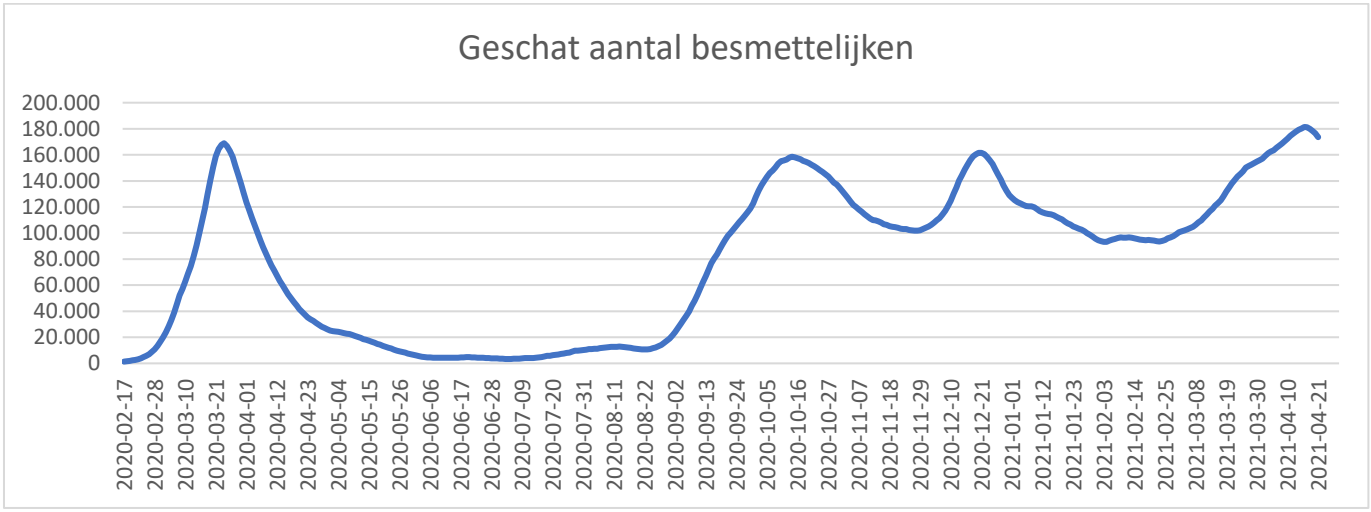


Figure 13: B1: covid-19 cases wide blue

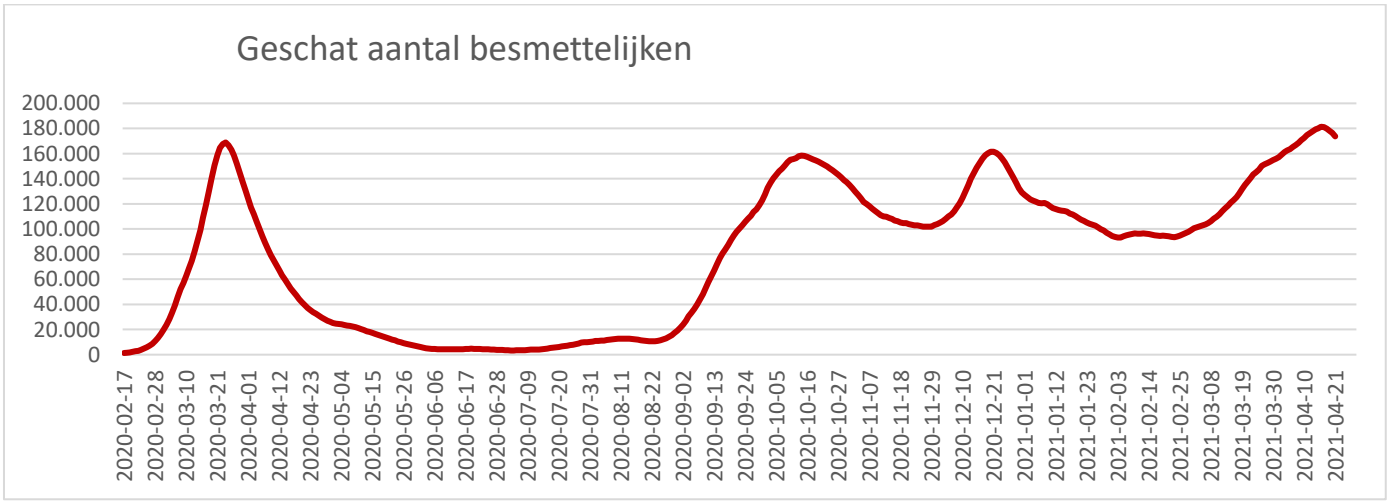


Figure 14: R1: covid-19 cases wide red

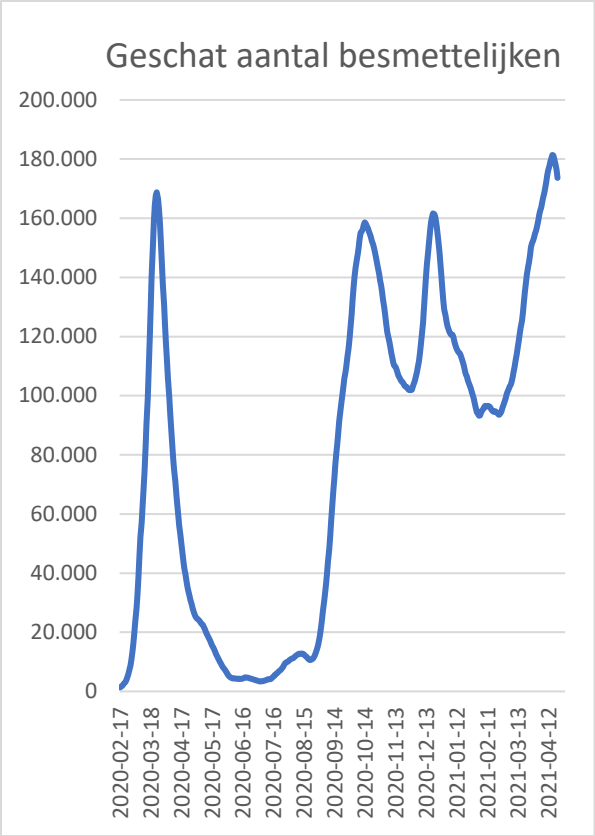


Figure 15: B2: covid-19 cases tall blue

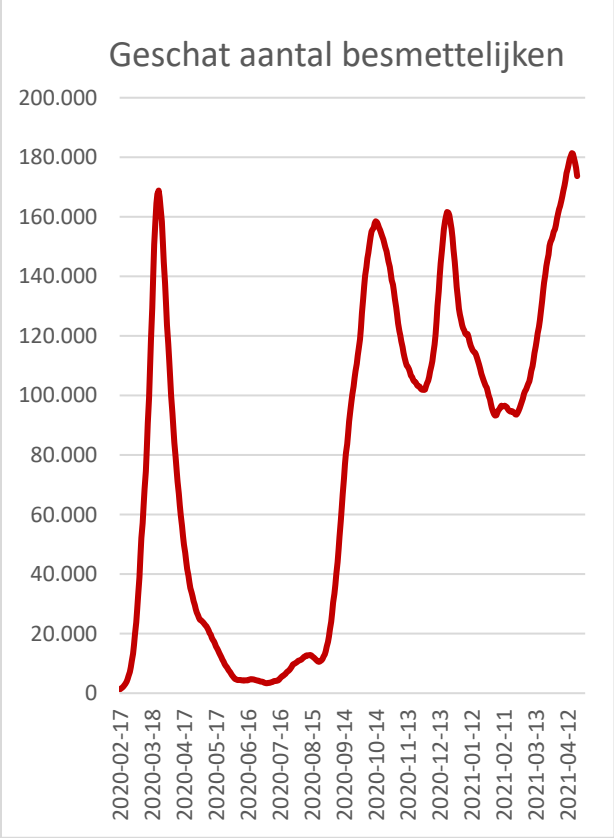


Figure 16: R2: covid-19 cases tall red

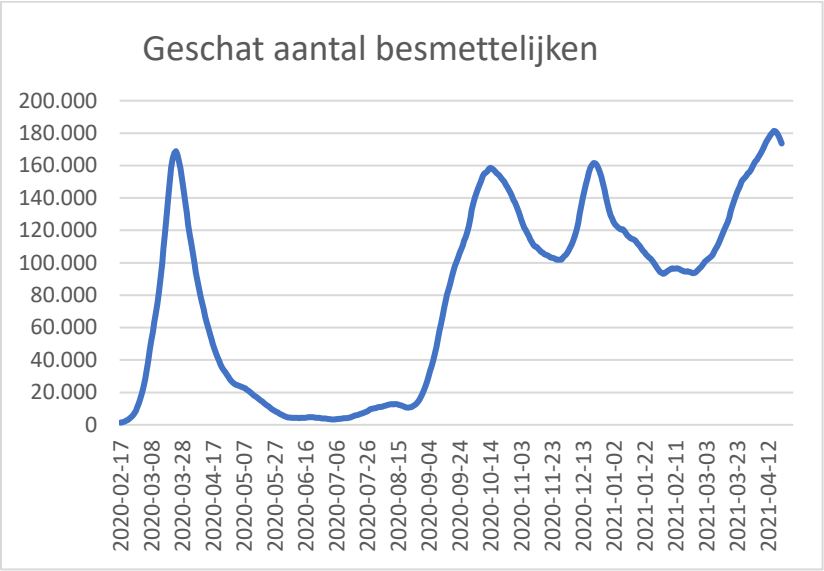


Figure 17: B3: covid-19 cases blue

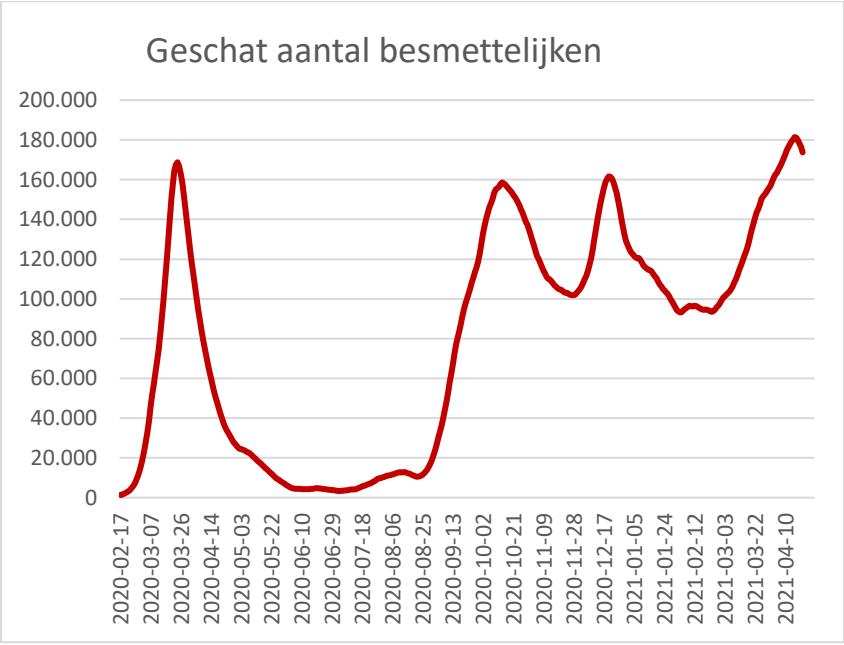


Figure 18: B3: covid-19 cases red

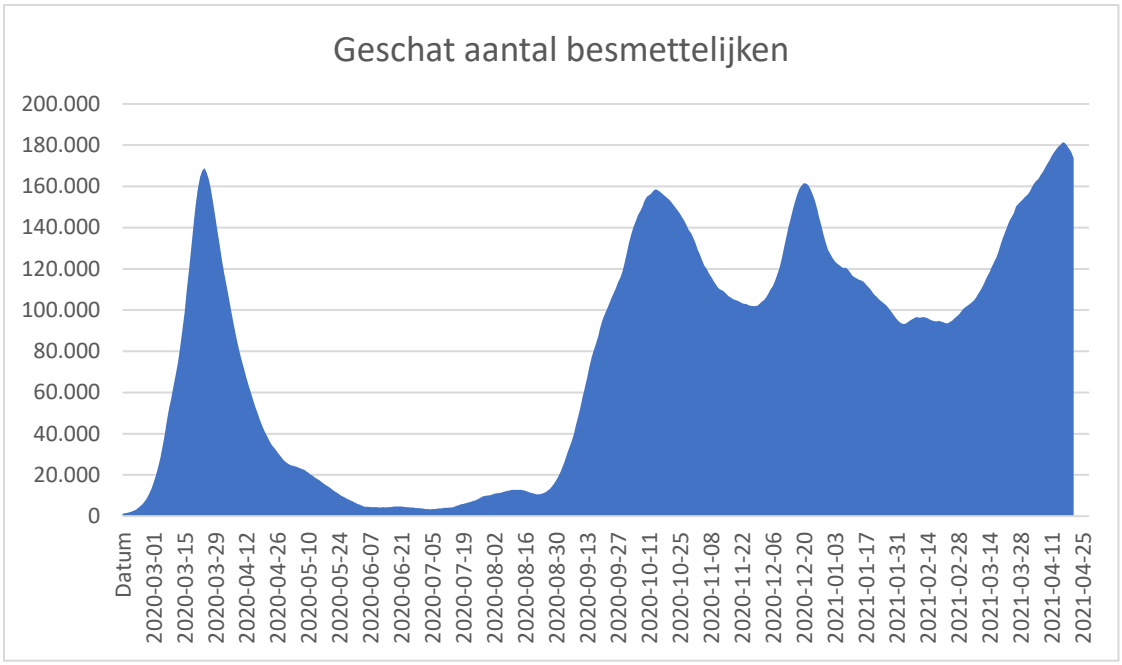


Figure 19: B4: covid-19 cases shaded blue

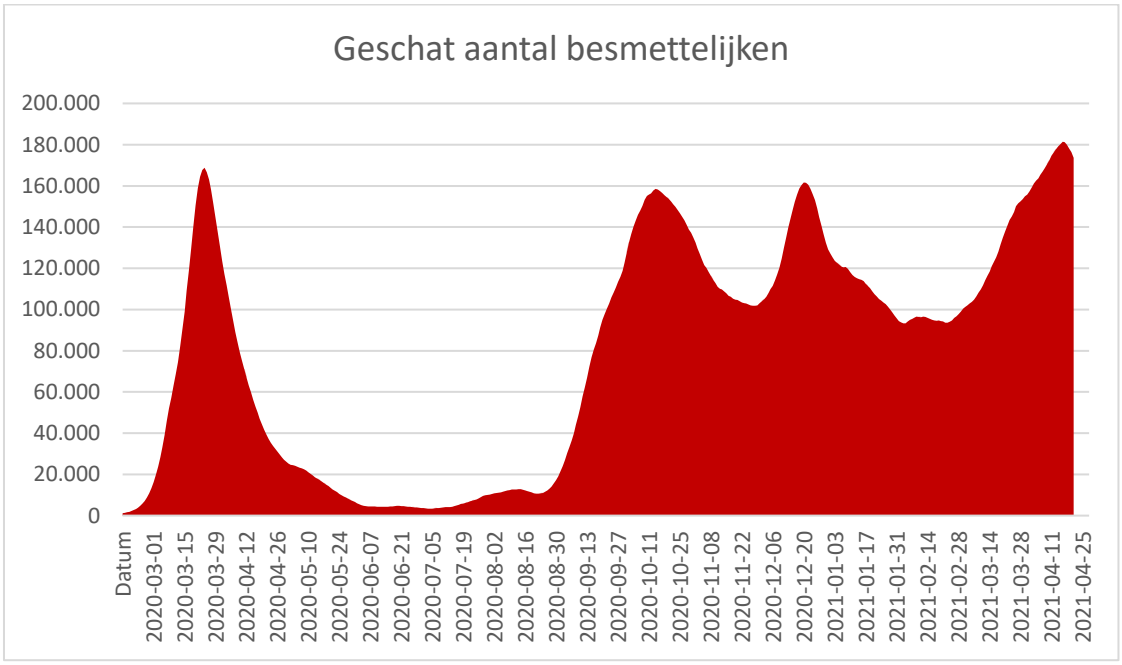


Figure 20:R4: covid-19 cases shaded red

Appendix C: Figures about how covid-19 cases are visualized in the Netherlands

**These graph visualizations were collected but were not used in the main body of the text.*

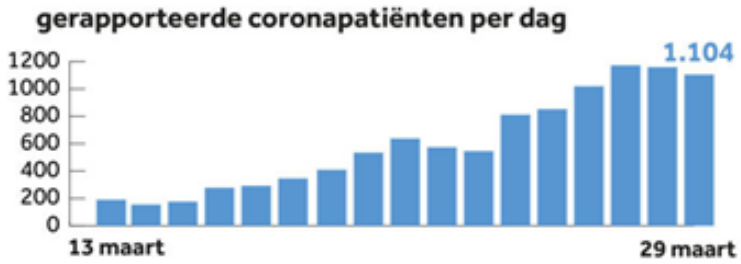


Figure 21: Covid in the Netherlands: numbers from March 29th (NOS, 2020)

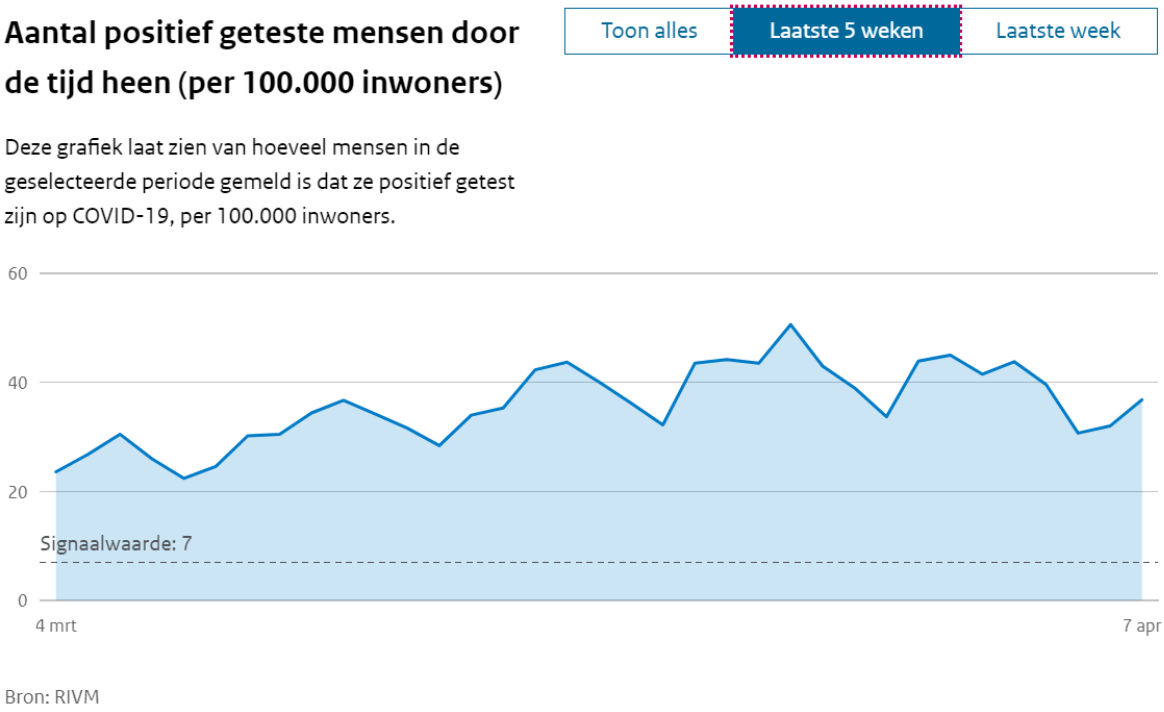


Figure 22: Amount of positive tested people through time (per 100.000 citizens) last five weeks RIVM (April 7th 2021)

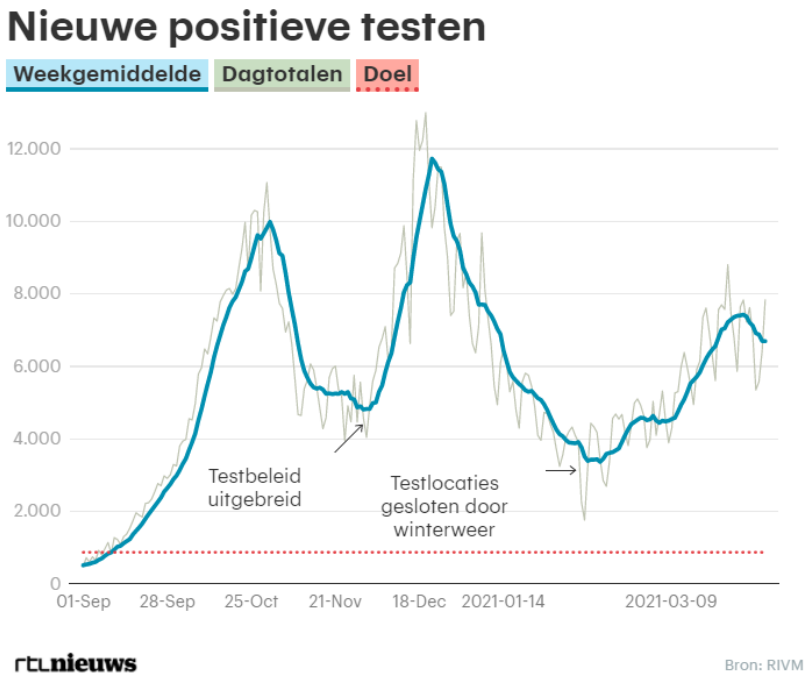


Figure 23: New positive tests week average RTL Nieuws (April 6th 2021)

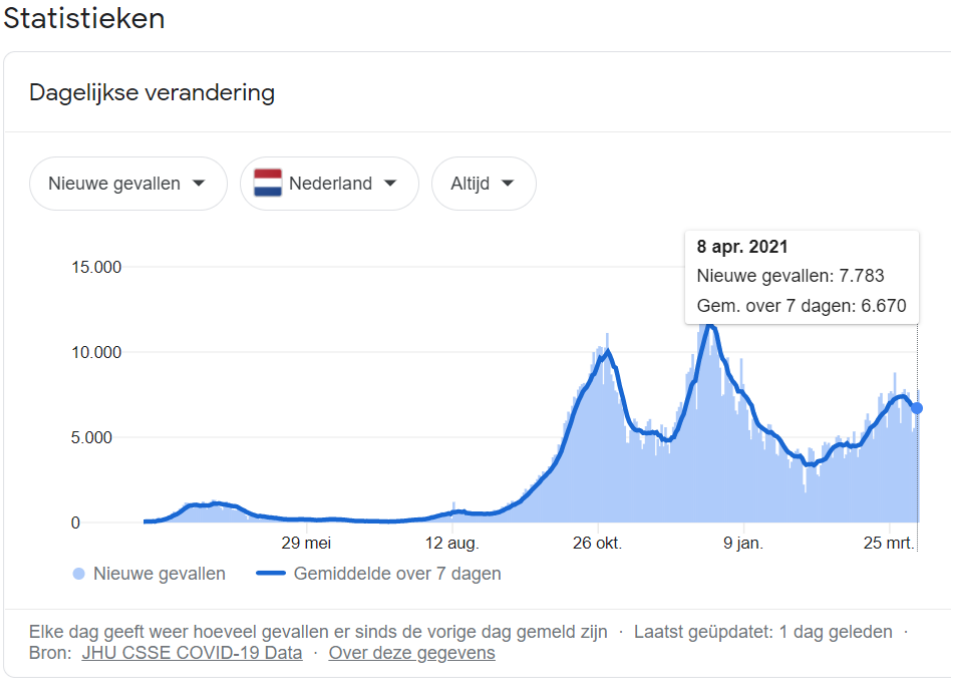


Figure 24: COVID-19 cases all-time Google (April 9th 2021)

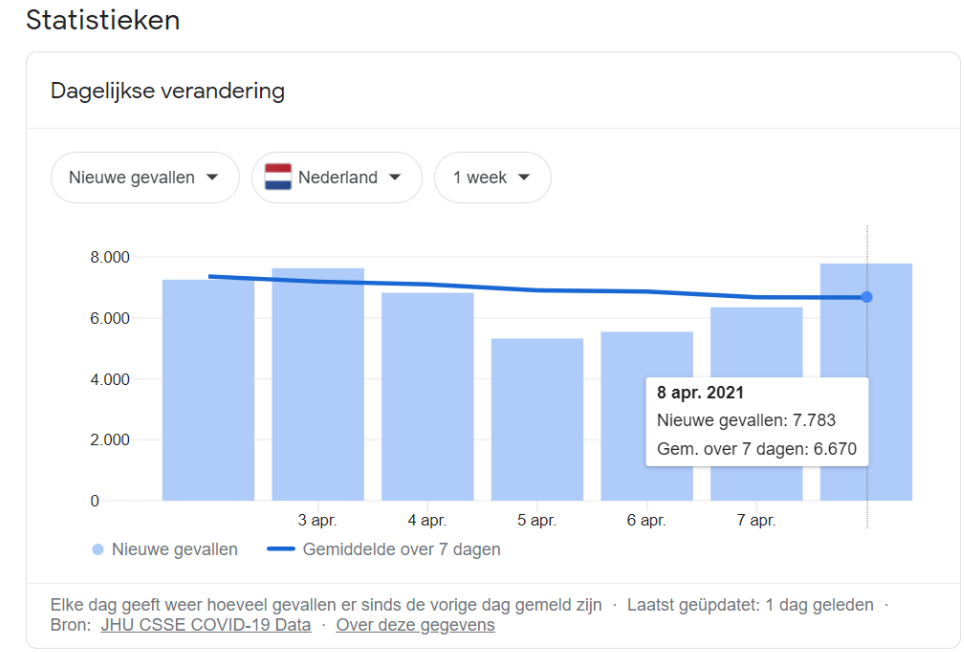


Figure 25: COVID-19 cases 1 week Google (9th of April 2021)

Figure 26: COVID-19 cases 30 days Google (9th of April 2021)

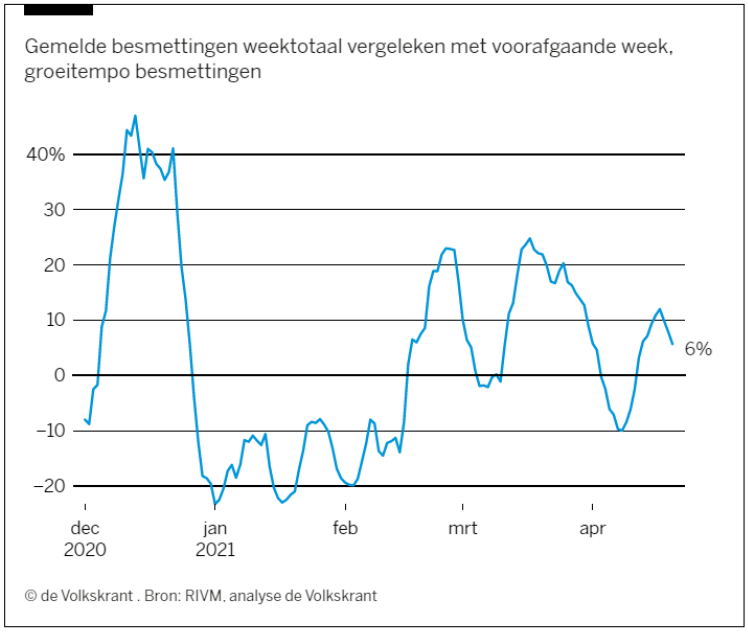


Figure 27: Growth rate infections Volkskrant (20th of April 2021)

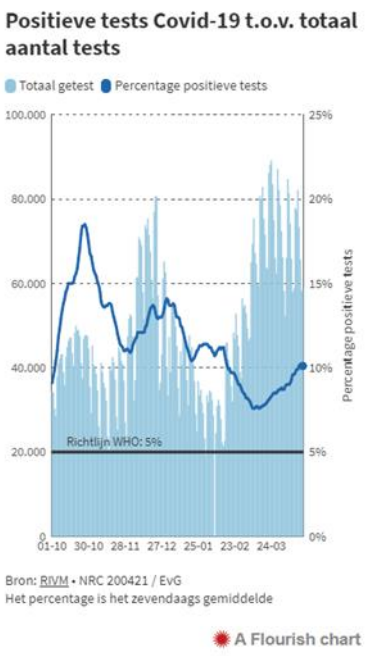


Figure 28: Positive percentage of COVID-19 tests NRC (21st of April 2021)

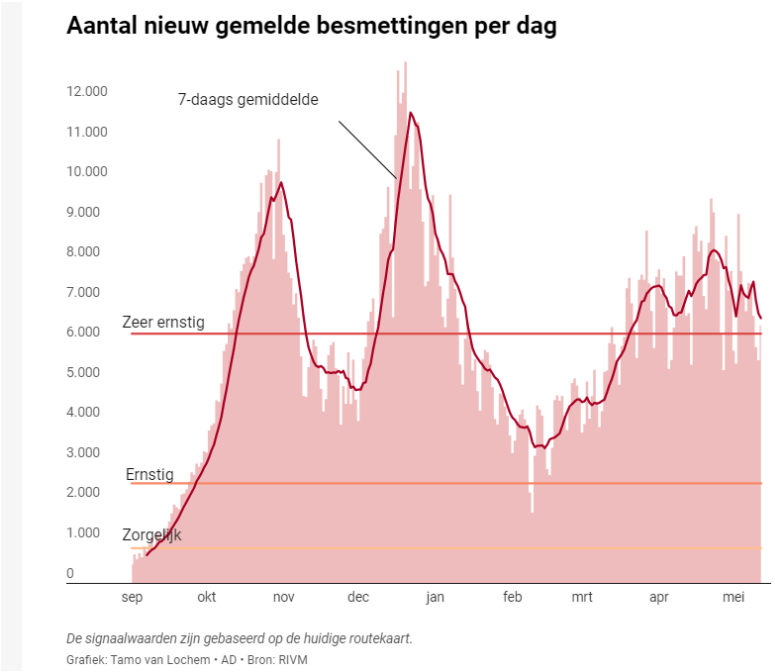


Figure 29: Amount of new covid-19 cases per day AD (8th of May)

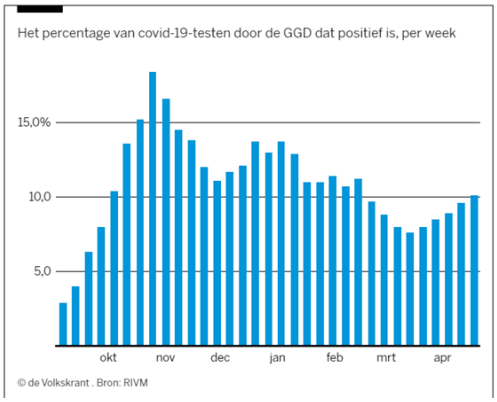


Figure 30: The percentage of COVID-19 tests that were positive, per week Volkskrant (20th of April 2021)