**Text Presentation Survey Data Analysis**

*Niet tekst, wel opmerkingen. Ik heb het volgende veranderd aan de functie:*

*Functie step 1 a en b*

* *Input used\_dates gemaakt zoals bij de andere functies ipv sewer\_data (although, I like that name.*
* *Ylim op 400,000 gezet, was 350,000, maar het totaal is aan het eind (november) meer dan dat en ik dacht dat we het limiet we hoger moesten zetten dan het totaal, omdat ik wel eens heb gehad dat R anders rare dingen doet om het in het frame te proppen.*
* *Kleur zoals besproken naar gifgroen gezet.*
* *Voor de titel figure 1a gezet (makkelijk verwijzen in de tekst). En bij b ook.*
* *Ik heb de namen van alle figuren veranderd zodat ze zo informatie mogelijk zijn en ook allemaal hetzelfde met alleen een kleine aanpassing die ertoe doet.*

*Wat ik me nog afvraag:*

* *Titel veranderen naar iets van “Grant Total…” zodat duidelijk is dat dit het totaal is van ooit gerapporteerde infecties in Nederland, ipv totaal gerapporteerde infecties op een dag alleen (soort van afgezonderd, op zichzelf staand)? Wanneer ja, dan zou ik bij de titel “per day” ook vervangen door “over time”.*
* *Willen we de schaal van de y-as verkleinen, als in, in duizendtallen? Of is dat eigenlijk al stap 2? Want nu staat er e+05 en dat vind ik lelijk.*
* *Waarom deden we het aantal reported ook alweer keer 100,000 / population. Waarom was total reported gedeeld door 100,000 niet genoeg? Ik ben dat deel van basisschool rekenen weer vergeten ben ik bang.*

**Nota bene naar mij en de rest:** gebruik infections ipv cases, want dat doen we op dit moment ook in alle functies.

Ik neem aan dat ergens hiervoor wordt genoemd dat we het totale aantal infecties van de RNA data proberen te schatten…

**Step 1 – The number of infections in The Netherlands**

Before taking on the task of predicting the total number of infections in the Netherlands on a given day from the RNA particles in the sewage, we have aggregated the data on the number of reported infections in a municipality on a given day to the total number of reported infections on that day in The Netherlands as a whole.

To achieve this, we made a function that takes the data as input. Then, for every day, the function first gets rid of any duplicate municipalities that are due to multiple sewage installation inputs for one municipality, and then sums the number of reported infections on a given day over all municipalities. The resulting output is visualized in Figure 1a, where you can see the total number of infections we have had over time in the Netherlands. The second wave is also clearly visibly, where you see the number of total infections increasing more rapidly in October compared to September.

Figure 1a

However, we were also interested in the daily fluctuations of the number of infections in the Netherlands. Therefore we altered the function that we had so that it would subtract the summed total number of infections of the previous from the summed number of infections of the day we wanted to know the increase for. After looping the function through every day in our dataset, the results are visualized in Figure 1b.

Figure 1b

**Step 2 – The number of infections per 100,000 inhabitants in The Netherlands**

Next, we visualized the number of cases per 100,000 inhabitants of The Netherlands. Again, we decided to do this for both the total number of infections we have had over time, as well as the daily differences. As you can see in Figures 2a and 2b, the shape of the distribution is the same, but the scale on the y-axis is different. This is expected, as we took the functions of the previous step and multiplied the total number of reported infections by the number of inhabitants of the municipality divided by 100,000.

Figure 2a

Figure 2b