

# Design 1: Population data

## Groep 4

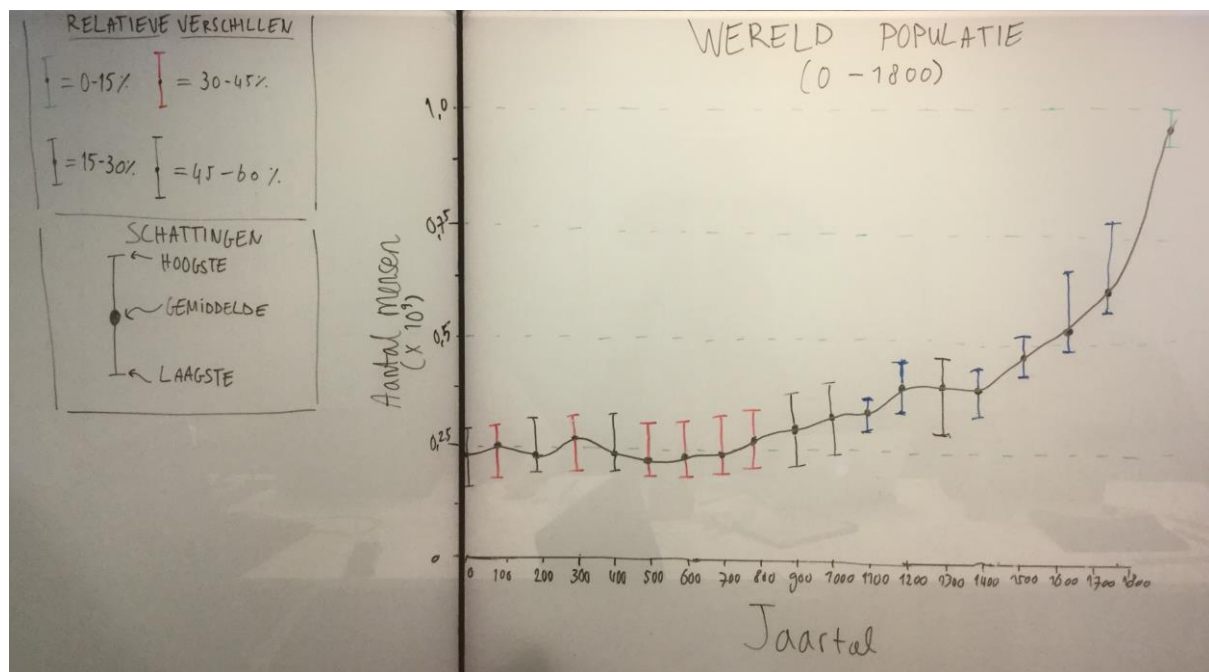
Sanne Meijering	10783709
Tim Molleman	10587306
Jelle Mul	11402148
Shan Li Nio	6222420
Misha Paauw	10054154

### Part 1

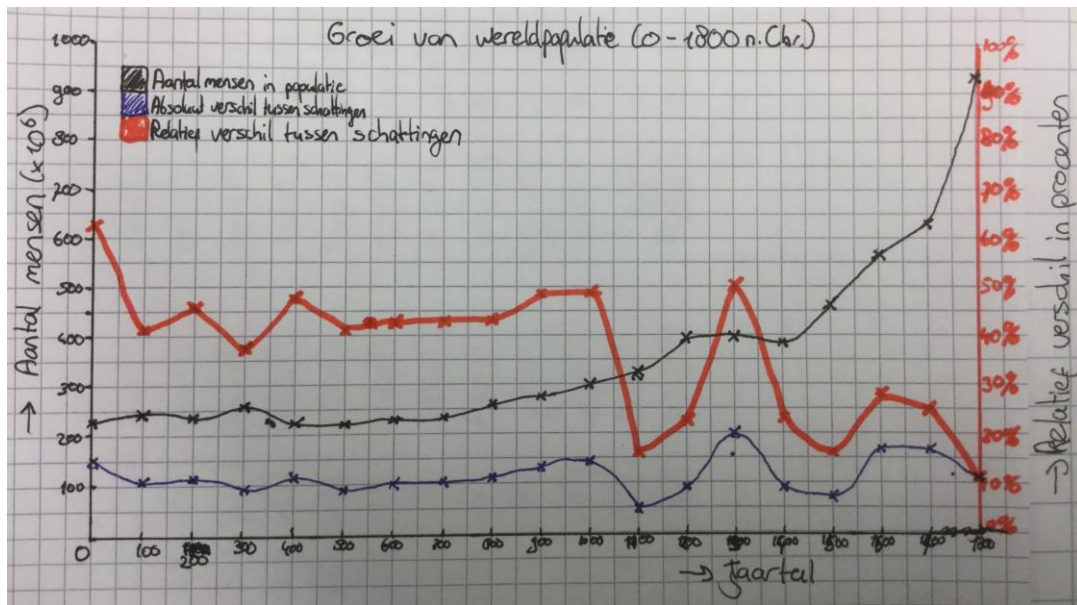
In de populatiedata is een exponentiële groei terug te vinden. Het absolute verschil tussen de schattingen lijkt geen trend te bevatten, maar het relatieve verschil heeft een dalende trend na het jaar 1000. Wat betreft de maximale verschillen, is het absolute verschil het hoogst vanaf het jaar 1800. Dit wordt verklaard door de vele male grote wereldpopulatie wat absoluut gezien een bredere spreiding geeft. De relatieve verschillen zijn het grootst in de periode voor het jaar 1000, wat wordt verklaard door het gebrek aan literatuur uit deze tijd.

Het is mogelijk om de onzekerheid en de data te plotten in één grafiek, hiervoor kan gebruik worden gemaakt van verschillende lijnen en assen, of door de minimale en maximale schatting aan te geven. Wanneer er sprake is van een exponentiële plot, zal lineaire interpolatie geen goede methode zijn voor het weergeven van de onzekerheid. Er wordt namelijk een overschatting gemaakt van de onzekerheid. Deze overschatting neemt toe, naarmate er verder in de grafiek wordt gekeken en de helling van de grafiek toeneemt.

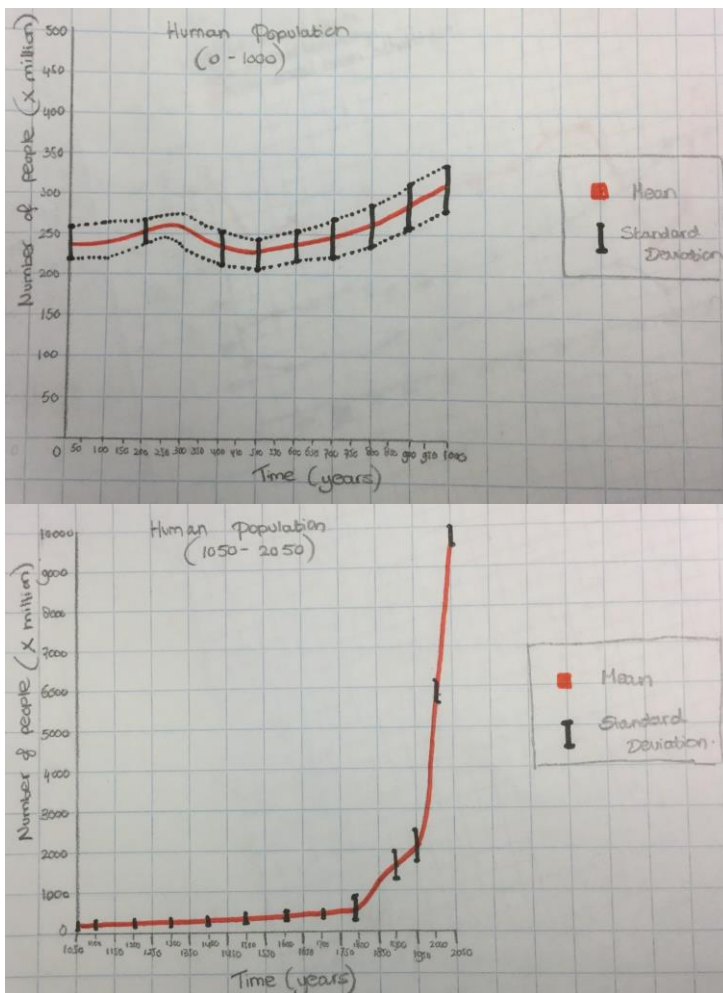
### Visualisaties



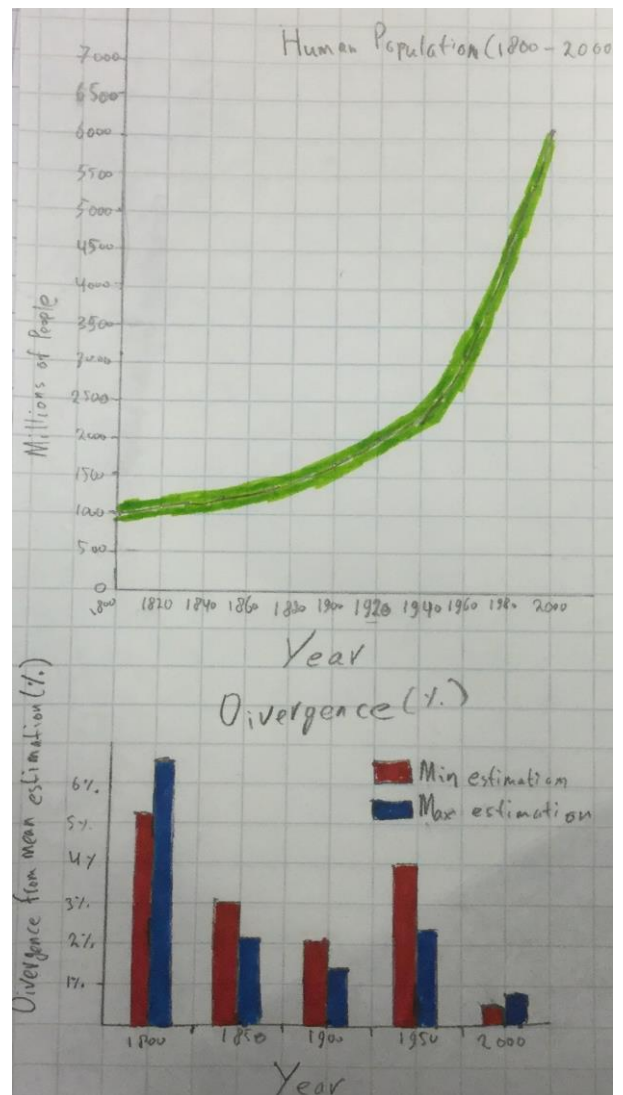
Figuur 1. De naar onze mening beste visualisatie. Visueel duidelijk, alle schattingen zijn meegenomen en berekend tot een gemiddelde. Verder bevat de grafiek zowel de relatieve (d.m.v. de verschillende kleuren) als de absolute verschillen (error balken) in data per jaar. De tijdsperiode die is gekozen (jaar 0 t/m jaar 1800) laat de meest interessante data en onzekerheid zien. Door het gebruik van horizontale rasterlijnen is de data ook verderop in de grafiek goed afleesbaar. Voordeel is dat er gebruik wordt gemaakt van slechts 2 assen, gevolg is dat de relatieve verschillen minder exact zijn weergegeven. Dit weegt echter op tegen de visuele onduidelijkheid die anders ontstaat.



Figuur 2. Visualisaties waarin alle informatie in 1 plot weergegeven, nadeel is echter de chaos die wordt gecreëerd door het gebruik van verschillende assen en 3 verschillende lijnen. Voordeel is dat alle punten tot op grote nauwkeurigheid kunnen worden afgelezen.



Figuur 4. Een combinatie van 2 plots voor 1 visualisatie. De data het gemiddelde van de schattingen, waar de error balken de standaarddeviatie aangeven. De bovenste plot is de periode 0 - 1000 en de onderste 1050-2000.



Figuur 3. Een combinatie van 2 plots voor 1 visualisatie, met in de bovenste plot de spreiding van de absolute schattingen tegen het jaartal. De tweede plot geeft de procentuele afwijking weer van de minimale en maximale schatting in de weergegeven jaren.