

Data Processing

Population Data

12-02-2016

Universiteit van Amsterdam

Minor Programmeren

Groep B1

Julia Jansen

Shannon Bakker

Tom Schoufour

Thom Mekelenkamp

1. Analyse

Trends in de data

- In de bijna gehele tabel stijgt het aantal mensen op de wereld.
- T/m 1500 stijgt de populatie gematigd, na 1500 neemt de groei per jaar sterk toe.
- Er zijn voor 1950 een aantal periodes waarin de wereldpopulatie krimpt.
- Vanaf 1900 begint de bevolking explosief te groeien
- Vanaf 1950 groeit de bevolking met ongeveer 100 miljoen per jaar.
- Vanaf 8 miljard mensen lijkt de groei te stabiliseren.

Verschillen in de data

- Voor 1850 zijn er grote verschillen tussen de schattingen van verschillende onderzoekspartijen.
- Sommige partijen geven slechts eens in de paar jaar een schatting, waar andere partijen ieder jaar een schatting uitbrengen.
- US, UN en Maddison lopen vrijwel gelijk op vanaf 1950. Maddison group ligt van 1950 tot 2003 onder de schattingen van de US, UN en Maddison. Na 2003 liggen de schattingen juist ver boven die van de andere onderzoeken.

Verschillen in schattingen (absoluut en relatief)

- Verschillen tussen de schattingen worden over de tijd een stuk groter. Maar naarmate de populatie groeit worden de relatieve verschillen ten opzichte van de populatie kleiner.
- Absoluut het grootste verschil:
 - Het verschil in de schatting in 2010 tussen Maddison en de USCB is absoluut het grootst met een verschil van ongeveer 900 miljoen.
- Relatief het grootste verschil:
 - Het verschil tussen HYDE en McEvedy & Jones in het jaar -4000 is relatief het grootst. De partijen verschillen in hun schattingen maar liefst 400% van elkaar.

Data representatie

Het is mogelijk om in één grafiek zowel de bevolkingsgroei als de relatieve verschillen tussen de verschillende schattingen weer te geven. Dat kan worden weergegeven door de schaal van de as aan de totale populatie over de tijd aan te passen. Wanneer dit toegepast wordt, wordt het relatieve verschil in de schattingen waarheidsgetrouw weergegeven. Het nadeel van deze vorm is wel dat de geschaalde as dermate lastig te lezen wordt dat het niet aan te raden is om deze vorm te gebruiken. Bovendien zal de trendlijn van de groei er dan niet exponentieel maar meer lineair uitzien, wat de groei minder waarheidsgetrouw weergeeft.

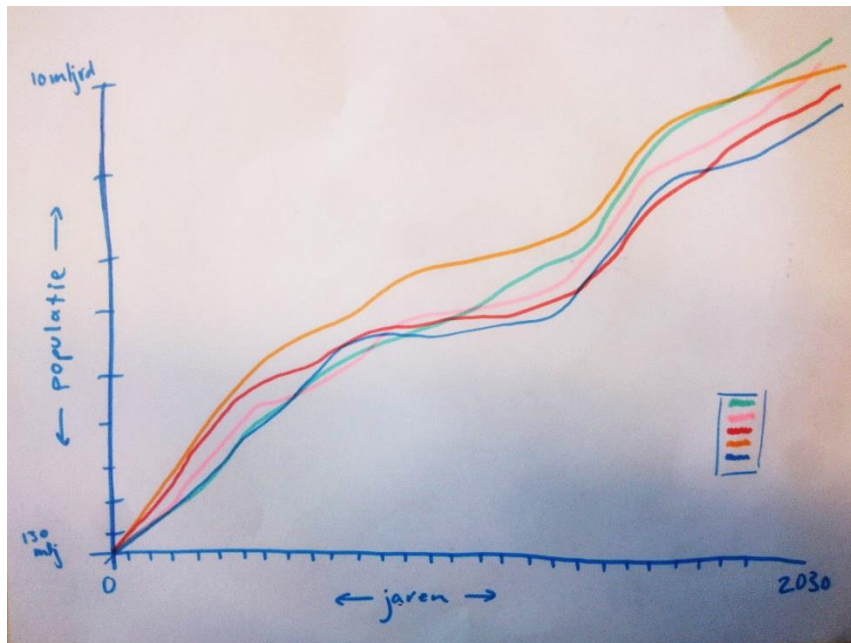
Lineaire interpolatie

Alleen als de data op de zelfde jaren is gebaseerd en de intervallen niet te groot zijn kan je lineaire interpolatie gebruiken. Als dit niet het geval is, is er sprake van een soort schijnzekerheid. Er wordt met lineaire interpolatie namelijk een lijn getrokken van schatting naar schatting, door punten waarvan helemaal geen schatting gedaan is. Als de schattingen ver uit elkaar liggen, zal tussen die schattingen toch data afleesbaar zijn uit de grafiek, terwijl er geen data gemeten is in die jaartallen.

Het zou wel mogelijk zijn lineaire voor Madison Group, Madison en de UN vanaf 1950 tot en met 2009.

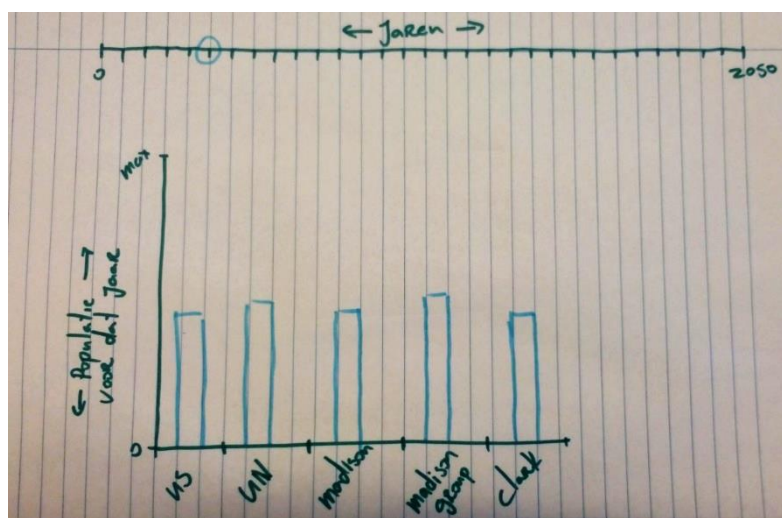
2. Schetsen

1. In de grafiek in afbeelding 1 is de y-as is een geschaalde as op basis van de populatiegrootte. Op de x-as staan de jaartallen van 0 t/m 2050. We hebben voor een geschaalde as gekozen, omdat je op deze manier de relatieve verschillen tussen de schattingen eerlijker kan zien. Als je dit niet doet dan lijken de verschillen in het begin heel klein, omdat ze absoluut klein zijn in vergelijking met latere jaren, terwijl ze relatief heel groot zijn. Iedere lijn in de grafiek geeft de data van één instituut weer.



Afbeelding 1

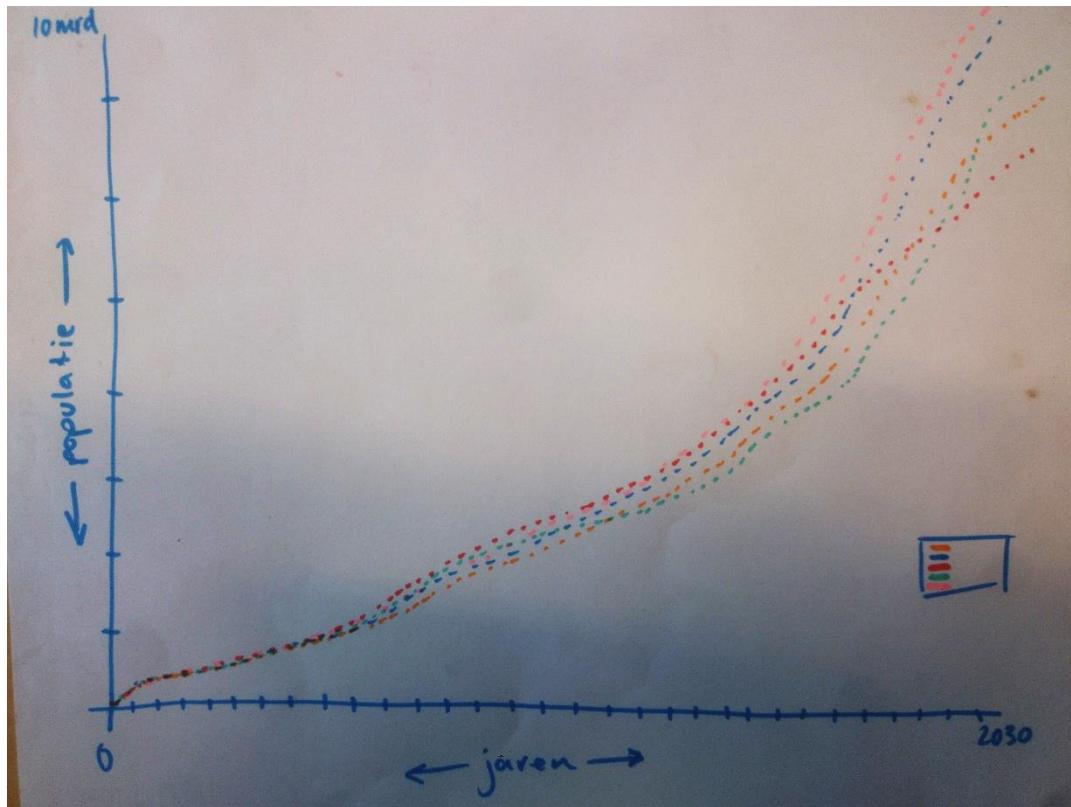
2. De grafiek in afbeelding 2 is een staafdiagram. Op de tijdlijn bovenaan de grafiek kan je kiezen van welk jaar je de gegevens wilt zien. Wanneer er op een jaar geklikt wordt krijg je alle data van dat jaar met op de x-as de verschillende instituten en op de y-as een schaal passend voor dat jaar. 'Passend' betekent dat de hoogste waarde van de schaal gelijk is aan de hoogste schatting van de populatie voor dat jaar. Op deze manier zijn de relatieve verschillen goed zichtbaar.



Afbeelding 2

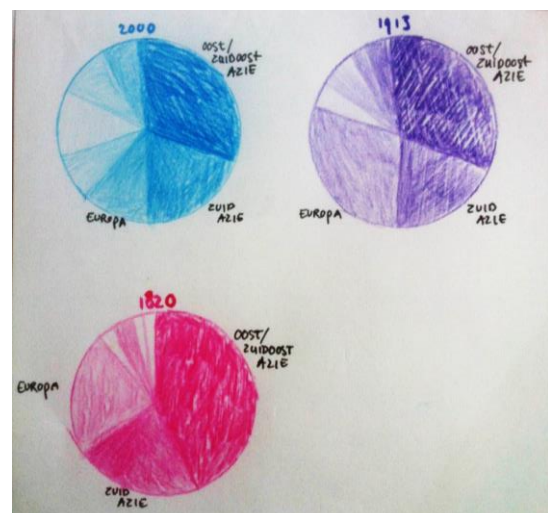
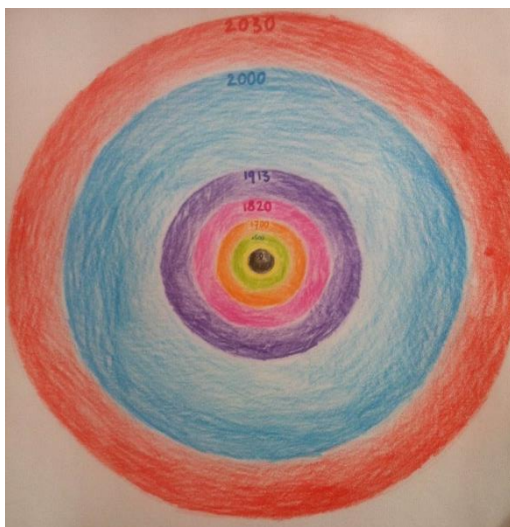
3. De grafiek in afbeelding 3 heeft een lineair geschaalde y- as, met daarop de populatiegrootte en een x-as waarop alle jaren staan. Het voordeel van een lineair geschaalde y-as is dat de groei van de bevolking duidelijk is weer te geven. Daarnaast is de y-as makkelijk af te lezen. De data wordt in deze grafiek weergegeven met stippen en ieder onderzoek heeft een eigen kleur. We hebben gekozen voor stippen omdat je zo duidelijk kan weergeven hoeveel datapunten ieder onderzoek heeft.

Afbeelding 3



4. Grafiek vier laat met behulp van bollen de populatiegrootte zien. Wanneer men op een bol klikt ziet men de populatieverdeling over de verschillende continenten. Deze grafiek is visueel aantrekkelijker dan een lijn- of staafdiagram. Je ziet direct dat de populatie groter is geworden, omdat de bol ook groter is geworden.

Afbeelding 4



3. Discussie

Welke grafische weergaves niet?

We bespreken eerst waarom we sommige grafieken niet gaan gebruiken. Ten eerste gebruiken we grafiek 1 niet. Het belangrijk nadeel van de eerste grafiek is dat je de groei niet goed kan zien, dit komt door de geschaalde as. Daarnaast trek je een lijn tussen verschillende datapunten terwijl het aantal datapunten per onderzoek verschilt.

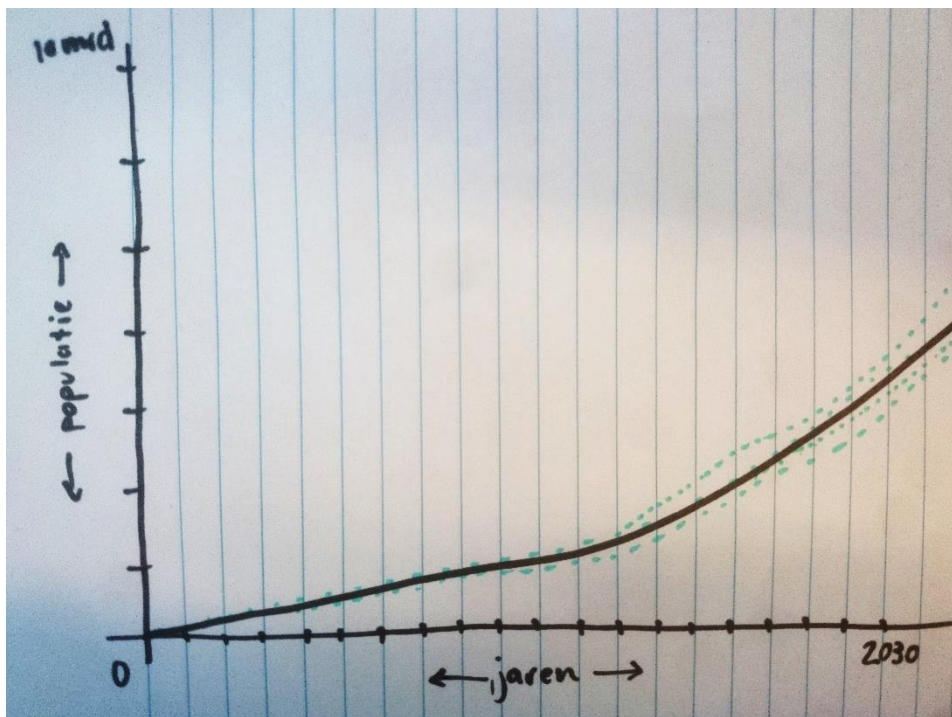
De derde grafiek gebruiken we ook niet. Dit komt doordat de stippen erg moeilijk leesbaar zijn. De oorzaak hiervoor is dat er erg veel datapunten zijn. Daarnaast zijn door de lineair geschaalde y-as de relatieve verschillen tussen de schattingen niet goed te zien aan het begin van de grafiek.

De vierde grafiek gebruiken we niet, omdat we hier niet goed de verschillen tussen de partijen in kunnen laten zien.

Welke grafische weergave wel?

We kiezen ervoor om de tweede grafiek wel te gebruiken. Het voordeel van deze grafiek is dat de verschillende voorspellingen van de partijen duidelijk zichtbaar zijn. Door de schaal steeds aan te passen is zowel het relatieve als het absolute verschil tussen de partijen duidelijk zichtbaar. Het nadeel van deze grafiek is wel dat het bijna niet te zien is hoe de populatie zich over meerdere jaren ontwikkelt. Om dit op te lossen kiezen we ervoor om de lezer eerst een lijngrafiek te laten zien. Op deze lijngrafiek staat er in het zwart een trendlijn voor de bevolkingsgroei en in het lichtgrijs staan de stippellijnen van de individuele onderzoeken weergegeven (zie afbeelding 5). Wanneer een lezer met zijn muis over één van deze stippellijnen gaat wordt de stippellijn uitgelicht en wordt weergegeven wie de bron is voor de schattingen in deze stippellijn.

Wanneer de lezer met zijn muis over de trendlijn gaat ziet men op de plek van de muis de staafdiagram voor dat jaar met alle verschillende onderzoeken (zie afbeelding 6). Deze staafdiagram ziet eruit als de tweede grafiek, alleen dan zonder lijn om een jaartal te kiezen. De lijngrafiek die erachter ligt verbleekt dan zozeer dat de staafdiagram goed te lezen is.



Afbeelding 5

Afbeelding 6

