



Centraal Bureau voor de Statistiek

Divisie Methoden en Kwaliteit
Sector Methodologie

*Postbus 24500
2490 HA Den Haag*

Visualisatieopdrachten CBS

Edwin de Jonge

Kennisgeving:

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

Projectnummer:

CBO-

BPA-nummer:

-CBO

Datum:

12-11-2010

VISUALISATIEOPDRACHTEN CBS

1. Inleiding

De kernactiviteit van het CBS is het verzamelen, bewerken, samenstellen en publiceren van statistische data. Van oudsher ligt in statistische bureaus zoals het CBS de focus op cijfers in tabelvorm. Tabellen zijn zeer nuttig, maar voor veel taken binnen het CBS niet de meest inzichtelijke of efficiënte methode.

Het onderzoekprogramma van het CBS bevat het onderdeel datavisualisatie. Het CBS haalt graag universitaire kennis in huis door bijvoorbeeld afstudeerstages Dit document bevat een aantal mogelijk opdrachten. Let wel dit is geen sluitende lijst, we staan open voor andere ideeën, zoals die in het onderzoeksprogramma passen en van potentieel van nut zijn voor het CBS.

2. Publicatie

2.1 Gebruik van animatie

Het gebruik van animatie voor het tonen van statistische resultaten is een vrij recent fenomeen. Bewegende grafieken/figuren kunnen het begrip van de data verhogen en het verhaal duidelijk maken dat verborgen was in de data. Helaas wordt vaak in de publicatie van statistische cijfers animatie toegevoegd om de data sexy te maken.

Welk gebruik van animatie effectief is en toegevoegde waarde heeft voor het tonen van statistische informatie. Dit kan eventueel door animatieprincipes zoals verwoord in *The Illusion of Life* (Disney Animation) te concretiseren voor statistische data. Een voorbeeld van dergelijke animatie is te vinden bij Gapminder (Human Development Trend 2005). Hierbij is het wenselijk dat de animatieprincipes concreet genoeg zijn, maar ook breed toepasbaar in bijvoorbeeld een visualisatie-toolkit.

2.2 Visualisatie toestand StatLine

StatLine (<http://statline.cbs.nl>) is de publieke database van het CBS. Hierin worden alle statistische cijfers van het CBS gepubliceerd in de vorm van multidimensionale tabellen (kubussen). De huidige StatLine database bevat meer dan 1500 van dergelijk tabellen waarvan een aantal zeer veel data bevatten (bijv. 100 miljoen cijfers). StatLine kent een aantal moeilijkheden voor gebruikers . Sommige tabellen zijn schaars (sparse) gevuld, waardoor

gebruikers data niet vinden. Data kunnen verspreid zijn over meerdere aanverwante tabellen/kubussen waardoor gebruikers data niet met elkaar kunnen verbinden en er kunnen inconsistenties tussen verschillende tabellen zitten.

De visualisatieopdracht is het ontwerpen van een visualisatie(-algoritme) om de vulling en samenhang van StatLine in beeld te brengen. Een uitgebreidere mogelijkheid is om potentiële inconsistenties in StatLine naar voren te brengen.

3. Dataverwerking

3.1 “Koppelen”

Het CBS maakt voor veel statistieken gebruik van meerdere databronnen. Één van de problemen bij het maken van een dergelijke statistiek is het combineren van de databronnen. Dit wordt in CBS jargon: “koppelen” genoemd. Vaak bevatten beide bronnen onvolledige informatie wat de koppeling bemoeilijkt, beschrijven beide bronnen een net andere populatie of net andere definities van variabelen. De opdracht is een visualisatie te ontwerpen die laat zien hoe goed databronnen gekoppeld kunnen worden, de grootte van de problemen laat zien en waar de problemen zich eventueel bevinden.

4. Analyse

Veel officiële statistieken zijn zogenaamde stroomgegevens waarbij de kwantitatieve variabele aangeeft hoeveel er stroomt tussen twee objecten. Men denke aan migratie, Nationale Rekeningen en leerlingen in het onderwijs. Nog te weinig worden deze gegevens getoond en geanalyseerd als stromen. Een dergelijke visualisatie toont duidelijker de verbanden en context van de gegevens aan dan de gebruikelijke tabel.

Doel is van afstudeeropdracht is om te kijken in hoeverre bestaande stroomvisualisatietechnieken toepasbaar zijn voor het visualiseren en analyseren van statistische data en eventueel de bestaande technieken aan te passen.