Kirsten de Wit  
6063098  
Programmeerproject Data Visualisatie

# Target

Gedurende de opleiding Psychobiologie en Neurobiologie heb ik veel over de hersenen en hersenaandoeningen geleerd. De hersenen maken deel uit van het centrale zenuwstelsel en er worden allerlei functies gereguleerd, zoals de hartslagfrequentie en de ademhaling. Daarnaast vinden de mentale processen zoals denken plaatst in de hersenen en ook wordt de motoriek getuurd vanuit de hersenen. De hersenen zijn belangrijk en complex en als er iets fout gaat in de hersenen kan dat grote gevolgen hebben. Bijna iedereen krijgt of zelf of in de omgeving met de hersenaandoening te maken. Er zijn heel veel verschillende hersenaandoeningen een voorbeeld is de ziekte van Alzheimer. Bij de zieke van Alzheimer

en bij deze aandoeningen zijn verschillende genen betrokken en naar de rol van deze genen wordt steeds meer onderzoek gedaan. Alleen het visueel maken van welke genen bij welke hersenaandoeningen betrokken zijn ,zover ik gehoord heb en kon vinden, niet gedaan. Door deze genen te visualiseren is het makkelijker te zien of dezelfde genen bij verschillende hersenaandoeningen betrokken zijn.   
Er is voor deze visualisatie gekozen om 5 hersenaandoeningen (Ziekte van Alzheimer, Ziekte van Parkinson, Ziekte van Huntington, Depressie en Epilepsie) te visualiseren. Daarnaast is er gekozen om de 20 meest relevante genen (op 29 mei 2015) op het genen gedeelte van PubMed (http://www.ncbi.nlm.nih.gov/gene/) te selecteren. Deze data is handmatig verzameld in een csv bestand, waarin de ziekte, de rank van relevantie, de naam van het gen, de positie, de link naar de informatie van het gen.

# Translate

# Design

Bij de eerste versie van de visualisatie werd er een scatterplot getoond waarbij de x-positie de positie op het chromosoom was en de y-positie was afhankelijk van welke ziekte het was. Hierbij viel al snel op dat dit niet ideaal was, omdat er tussen sommige punten veel witruimte zat en sommige punten zaten juist heel dicht op elkaar en waren nauwelijks te onderscheiden van elkaar. Vervolgens werd er gebruik gemaakt van de Fisheye Distortion (http://bost.ocks.org/mike/fisheye/) om er voor te zorgen dat de punten die dicht op elkaar lagen van elkaar te onderscheiden waren. De Fisheye zorgde er wel voor dat de punten groter werden wanneer je er met de Fisheye stond, maar het was nog steeds makkelijk om punten die vlakbij elkaar lagen van elkaar te onderscheiden. Er is daarom uiteindelijk gekozen om gebruikt te maken van de Cartesian Distortion. Door hier gebruik van te maken kan de breedte tussen de punten op de x-as veranderd worden, waardoor de punten (in het gedeelte waar de muis staat) verder uit elkaar komen te liggen. Hierdoor kan er makkelijker onderscheid gemaakt worden tussen de verschillende genen. Er moet gezegd worden dat voor enkele punten het onderscheid maken moeilijk blijft, doordat deze genen zo dicht na elkaar liggen op het chromosoom.  
Om onderscheid te kunnen maken tussen de 5 verschillende gekozen hersenaandoeningen, is er gekozen om deze hersenaandoeningen verschillende kleuren te geven. De kleuren zijn gekozen doormiddel van Colorbrewer 2.0 (http://colorbrewer2.org). Doordat er 5 kleuren nodig voor deze visualisatie is de visualisatie helaas niet kleurenblind vriendelijk.   
Aan de hand van Bertin’s karakteristieken van visualisatie zullen de eigenschappen en beslissingen die bij deze visualisatie een rol hebben gespeeld besproken worden. De visualisatie heeft de eigenschap positie, welke selectief is omdat de positie op de x-as onderscheiden kan worden. Daarnaast zit er een volgorde in de punten, maar alleen voor de rechts-links oriëntatie.   
Voor het aangeven van hoe relevant een gen was bij een bepaalde ziekte hebben verschillende karakteristieken van Bertin een rol gespeeld. In eerste instantie is er gekozen om de relevantie alleen aan te geven doormiddel van de specifieke kleur van donker naar licht te laten gaan. Aangezien het aantal kleuren dat onderscheiden (maximaal 10) kan worden beperkt is volgens Bertin’s karakteristieken was alleen het verschil in kleur niet voldoende. Daarnaast kwam bij dat kleuren niet een volgorde kunnen aangegeven. Aan de andere kant kon per kleur wel een verschil in waarde onderscheiden worden (donder naar licht) en dit zou wel een volgorder kunnen aangeven, maar ook hier kunnen maximaal 10 waardes onderscheiden worden. Er is daarom uiteindelijk gekozen om gebruik te maken van de combinatie van kleur en de grote van het ovaal. De grote van het ovaal is selectief, omdat alleen de grote van het ovaal kan er voor zorgen dat het onderscheiden kan worden van de rest. Daarnaast heeft het een volgorde en kunnen er maximaal 20 verschillen in grote van de ovaal onderscheiden worden.   
Om duidelijk weer te geven dat dezelfde genen bij verschillende hersenaandoeningen betrokken zijn, is er een filter functie ingebouwd. Hierbij wordt een bepaalde ziekte geselecteerd en dan worden bij de andere ziektes alleen de getoond die bij de geselecteerde ziekte betrokken zijn. Er is ook een tooltip toegevoegd met informatie, hierin staat de naam van het gen, de positie van de relevantie en de link naar de website met informatie over het gen. Om er voor te zorgen dat er op deze link geklikt kan worden is er een vertraging 5000ms ingebouwd. Daarnaast is de positie van de tooltip afhankelijk van de positie op het scherm, omdat deze anders buiten het scherm zal vallen.

# Implement

# Validate

Carpendale, M. S. T. "Considering visual variables as a basis for information visualisation." (2003).