

Naam: Kirsten de Wit

Studentnummer: 6063098

Email: kirstendewit91@gmail.com

Datum: 26 juni 2015

Programmeerproject Data Visualisatie

Inleiding

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 plaats 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 dit grote gevolgen hebben. Bijna iedereen krijgt of zelf of in de omgeving met de hersenaandoening te maken. Bij een hersenaandoening heeft iemand een beschadiging of een ziekte aan de hersenen. Wat er kan gebeuren bij een hersenaandoening is dat iemand dingen vergeet, of niet meer goed kan bewegen, of neerslachtig is, of niet meer goed kan spreken. Er zijn veel verschillende hersenaandoeningen 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, wordt 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 en de link naar de informatie van het gen staan.

Translatie

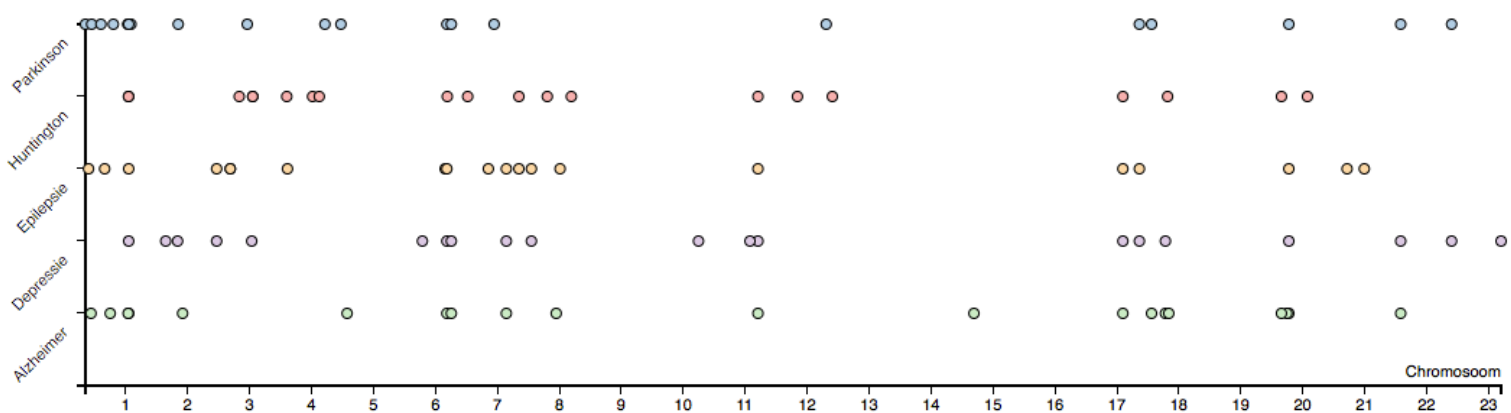
Om de data weer te geven werd gekozen voor een scatterplot, omdat hierin de relatie tussen 2 variabelen weergegeven kan worden. In dit geval is dat de relatie tussen de hersenaandoeningen en de genen. Een andere beslissing die genomen moest worden, was welke variabelen er op de assen moesten staan. Een optie was om de alle genen op één van de assen weer te geven, maar aangezien dit een maximum van 100 waardes (5 hersenaandoeningen * 20 relevante genen) kon opleveren was dit geen goede optie. Er is daarom gekozen om de 23 chromosomen op de x-as te zetten en de 5 hersenaandoeningen op de y-as. De keuzes op de variabelen op de x-as of y-as te zetten is gemaakt op basis van de hoeveelheid waardes de variabele had. Door de variabele met de meeste waardes op de x-as te zetten, kan er optimaal van de breedte van het scherm gebruik gemaakt worden.

Om de chromosoompositie van de genen te kunnen weergeven moest er een transformatie gedaan worden. De chromosoompositie wordt normaal gesproken als volgt weergegeven: 19q13.2 (APOE).

Om dit naar een getal om te rekenen is er gekeken bij welk basenpaar het gen begint en gekeken het hoeveelste basenpaar dit van het gehele chromosoom is. Door dit voor alle genen te doen, konden de genen op basis van de chromosoompositie weergegeven worden.

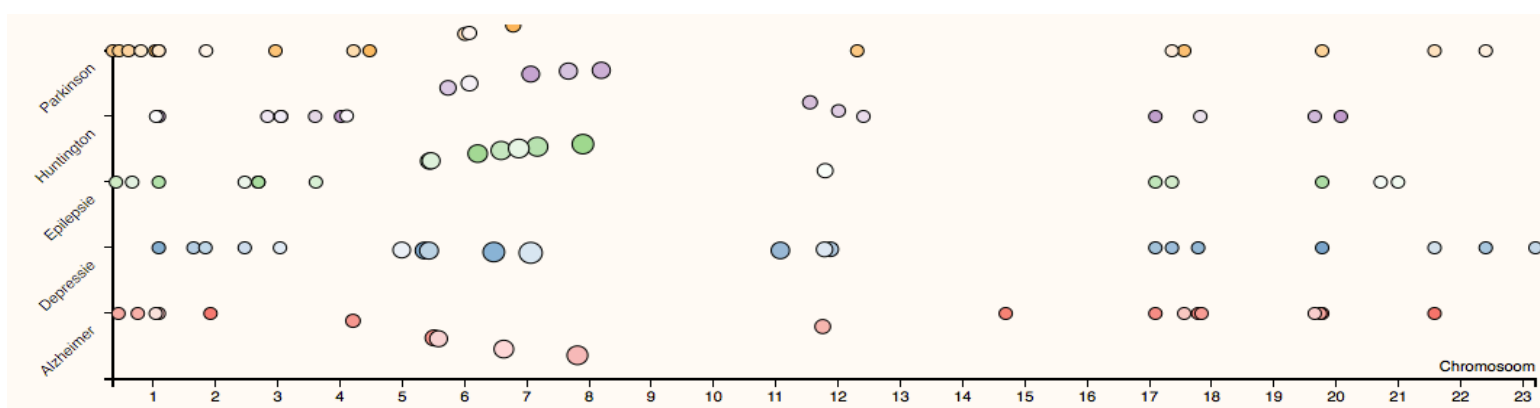
Design

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 (figuur 1).



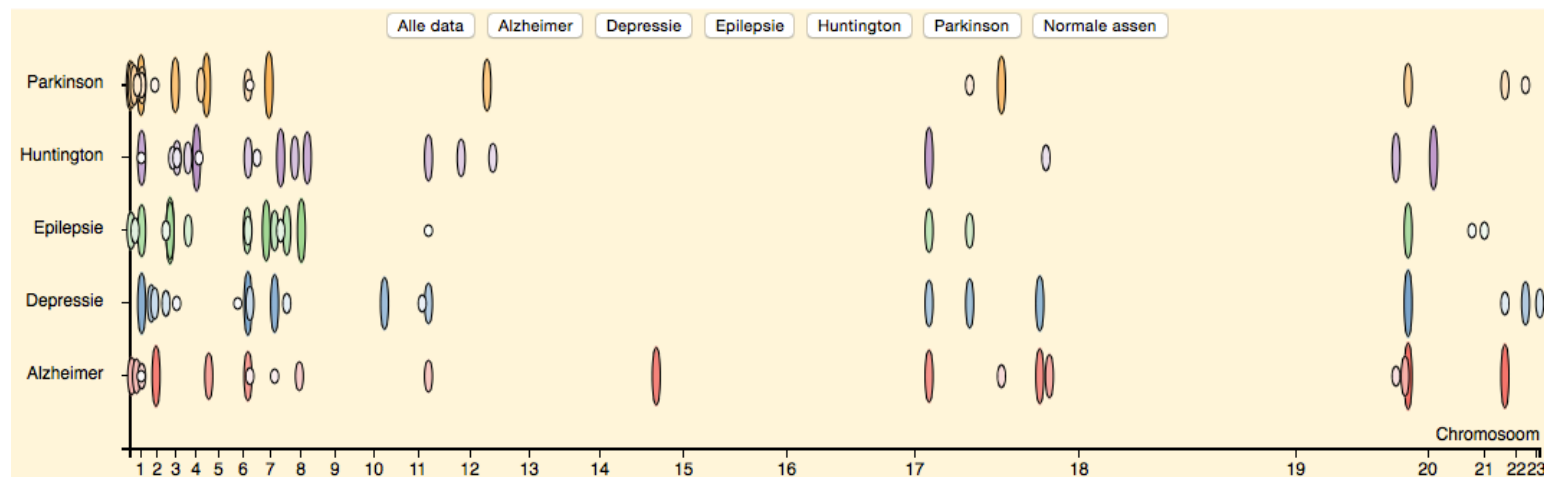
Figuur 1: De eerste scatterplot

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 niet makkelijk om punten die vlakbij elkaar lagen van elkaar te onderscheiden (figuur 2).



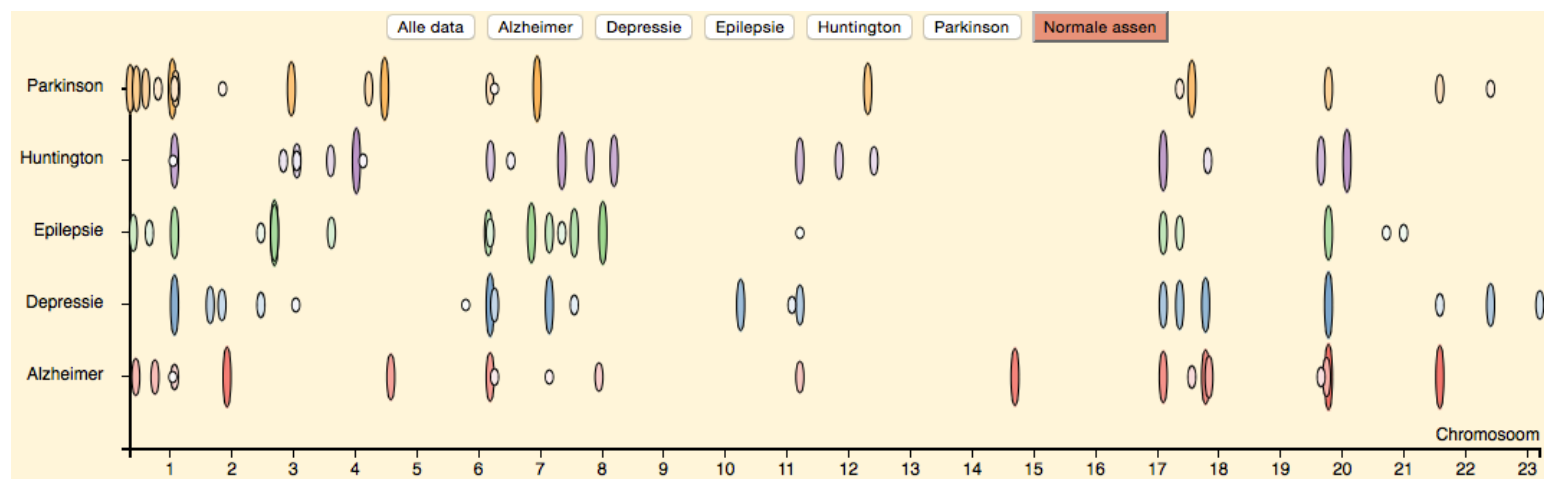
Figuur 2: Scatterplot met Fisheye

Er is daarom uiteindelijk gekozen om gebruikt te maken van de Cartesian Distortion Fisheye. 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 (figuur 3).



Figuur 3: Scatterplot met beweegbare assen

Er is naast de scatterplot met beweegbare assen ook nog de optie aan de visualisatie toegevoegd om te kiezen voor een grafiek met statische assen om bijvoorbeeld makkelijker de informatie die in de tooltip staat de bekijken (figuur 4).



Figuur 4: Scatterplot met statische assen

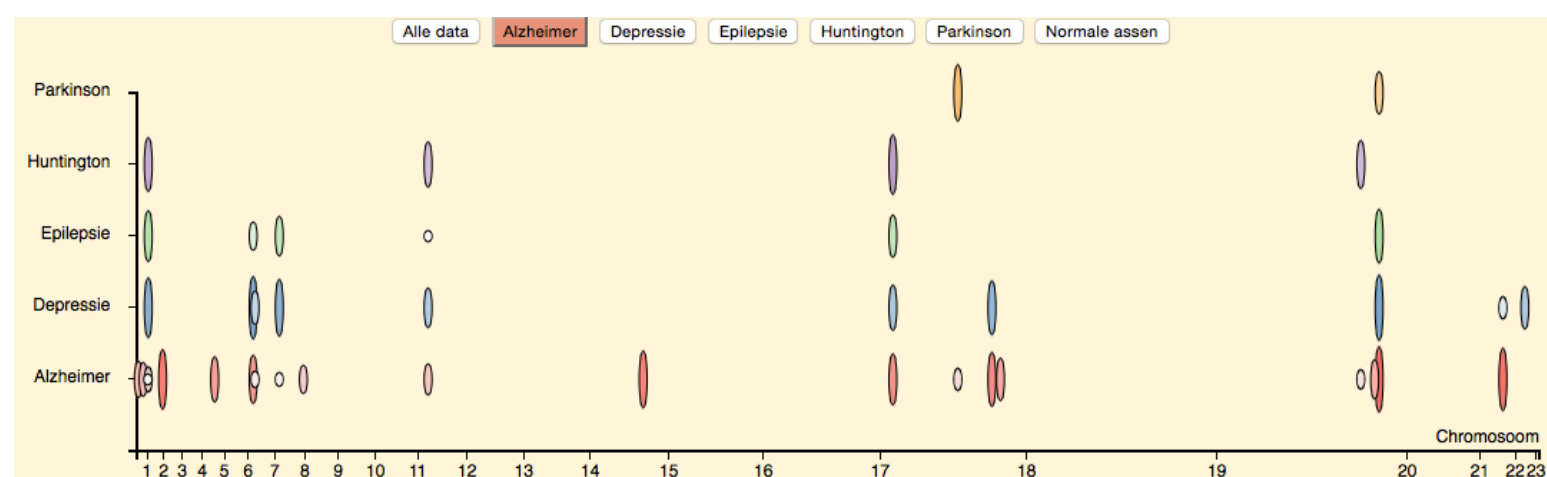
Om onderscheid te kunnen maken tussen de 5 verschillende gekozen hersenaandoeningen, is er gekozen om deze hersenaandoeningen verschillende kleuren te geven. De kleuren zijn gekozen doormiddel van Colorbrewer 2.0 (<http://colorbrewer2.org>). Doordat er 5 kleuren nodig voor deze

visualisatie is de visualisatie helaas niet kleurenblind vriendelijk.

Aan de hand van Bertin's (Carpendale, 2003) 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 er bij dat kleuren niet een volgorde kunnen aangegeven. Aan de andere kant kon per kleur wel een verschil in waarde onderscheiden worden (donker naar licht) en dit zou wel een volgorde 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 grootte van het ovaal. De grootte van het ovaal is selectief, omdat alleen de grootte 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 grootte van het 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 (figuur 5).



Figuur 5: Scatterplot met gefilterde data

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.

Implementatie

Voor deze visualisatie werden twee plugins gebruikt, namelijk D3 en D3 fisheye. In de code zitten 2 functies die allebei er voor zorgen dat de grafiek getekend wordt. De eerste (graph1) tekent een grafiek met de beweegbare as en de tweede (graph2) tekent een statische grafiek. Bij de grafiek met beweegbare as zit ook een filter functie. Deze filter functie werkt doormiddel van een key, die wordt meegeven aan “graph1” wanneer op de button die bij die ziekte hoort wordt gedrukt.

Evaluatie

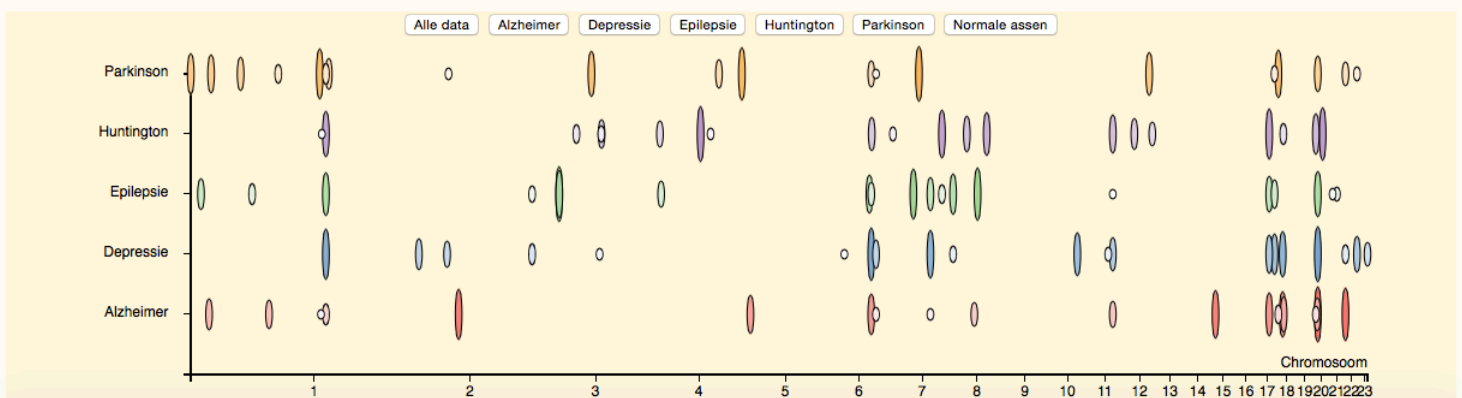
Over het eindresultaat ben ik tevreden en komt overeen met het in begin voorgestelde design, maar er hadden wel dingen anders gekund (figuur 6 en 7).

Hersenaandoeningen en Genen

Tijdens mijn opleiding Psychobiologie en Neurobiologie heb ik veel verschillende hersenaandoeningen voorbij zien komen. De laatste jaren wordt er, mede door de huidige technieken, veel meer gekeken naar de genen die betrokken zijn bij hersenaandoeningen. Het viel mij op dat de genen die betrokken zijn bij verschillende hersenaandoeningen nooit echt met elkaar vergeleken worden. Ook bij het zoeken op internet kon ik geen vergelijking vinden van genen die bij verschillende hersenaandoeningen. Wel is deze data te vinden op het genen gedeelte van PubMed (<http://www.ncbi.nlm.nih.gov/gene/>). Er zullen voor dit project 5 hersenaandoeningen (Ziekte van Alzheimer, Ziekte van Parkinson, Ziekte van Huntington, Depressie en Epilepsie) worden gekozen en van die aandoeningen worden de 20 meest relevante genen gekozen. Door het visueel maken hiervan kan er gekeken worden of verschillende genen betrokken zijn bij dezelfde hersenaandoeningen.

De visualisatie hieronder heeft op de x-as de 23 chromosomen en op de y-as de 5 verschillende hersenaandoeningen staan. De positie op de x-as geeft aan waar het gen zich op het chromosoom bevindt. Het verschil in kleur en grootte van het ovaal geven de relevantie aan, des te donkerder en groter des te relevanter. Daarnaast zal wanneer de muis op het ovaal wordt gehouden de informatie van het gen getoond worden, de rank in relevantie en de link naar de gen informatie op PubMed.

Doormiddel van de buttons kan een specifieke hersenaandoening geselecteerd worden, daarnaast zullen van de andere hersenaandoeningen de genen getoond worden die ook bij de geselecteerde hersenaandoening betrokken zijn.



Figuur 6: Het eindresultaat van de visualisatie

Ziekte van Alzheimer

Ziekte: De ziekte van Alzheimer (kortweg Alzheimer) zorgt voor 60% tot 70% van dementiegevallen. Het is een chronische neurodegeneratieve aandoening, waarbij het meest herkenbare symptoom het verlies van geheugen is. De oorzaak van Alzheimer is nog steeds niet bekend, maar het grootste risico komt vanuit de genen en daarnaast zijn andere risicofactoren depressie en hersenbeschadigingen. Alzheimer wordt geassocieerd met plaques en tangles die zich opbouwen in de hersenen. Onderzoek heeft aangetoond dat ouderen met een hoge opleiding minder kans hebben op het ontwikkelen van Alzheimer.

Diagnose: De diagnose wordt gedaan op basis van de Mini-mental state examination.

Behandeling: Op dit moment is er nog geen behandeling voor Alzheimer. Gemiddeld 6 jaar na de diagnose overlijden de mensen, meestal aan een longontsteking of uitdroging.

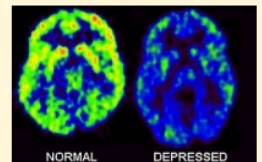


Depressie

Ziekte: Depressie is een psychische aandoening die zich kenmerkt door constante neerslachtigheid wat gepaard gaat met een laag zelfbeeld en verlies van levenslust. Bij het ontstaan van een depressie spelen verschillende factoren een rol; biologische, psychologische en sociale factoren.

Diagnose: De diagnose wordt gemaakt op basis van ervaringen van degenen zelf en de omgeving en de "mental status examination". Daarnaast wordt er gebruik gemaakt van DSM-IV TR, in dit handboek staat beschreven aan welke criteria voldaan moeten worden. De eerste keer dat iemand een depressie heeft is gemiddeld tussen de 20 en 30 jaar.

Behandeling: Er kunnen antidepressiva voorgeschreven worden, maar lijken het effectiefst voor mensen met een zware depressie te zijn. Daarnaast bestaat de behandeling vaak uit therapie.



Figuur 7: Eindresultaat van een deel van tekst over de hersenaandoeningen

Ten eerste de manier van filteren is niet helemaal netjes. Om te kunnen filteren zijn er extra variabelen aan het databestand toegevoegd, het was natuurlijk beter geweest om dit bijvoorbeeld via de DOM te kunnen selecteren. Een andere en wellicht betere optie om duidelijk te maken dat een bepaald gen ook bij andere ziektes betrokken is, is ervoor te zorgen dat die andere genen van de andere hersenaandoeningen gehighlight worden. Op deze manier zo dit ook toegepast kunnen worden bij de grafiek met de vaste assen.

Met het toevoegen van de beweegbare as, had ik gehoopt te kunnen oplossen dat punten die dicht bij elkaar liggen te kunnen onderscheiden. Dit is grotendeels gelukt, maar punten die heel dicht bij elkaar liggen, zijn nog steeds niet heel makkelijk van elkaar te onderscheiden. Wellicht zou dit nog opgelost kunnen worden door een zoomvak vast te maken aan de muis. Door dit te doen kan er goed ingezoomd worden op een specifiek gebied en hopelijk kunnen de punten dan wel onderscheiden worden.

Referenties

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

<http://colorbrewer2.org>

<http://bost.ocks.org/mike/fisheye/>

<http://www.ncbi.nlm.nih.gov/gene/>