

# *Job Market Tracker*, een design-critique.

Mats Willemsen

14 December 2015

## 1 Introductie

Het is voor veel Amerikaanse presidentskandidaten een belangrijk speerpunt geworden om te pleiten voor een goede positie van de Amerikaanse burger op de arbeidsmarkt. Deze positie is in verval geraakt als gevolg van een langdurige mondiale economische recessie, ook wel de *Great Recession* genoemd. De *Great Recession* is officieel beëindigd, maar de verslechterde economische omstandigheden hebben langdurige negatieve effecten op de gehele arbeidsmarkt.

Veel, vooral traditionele, media proberen de Amerikaanse burger op een overzichtelijke te informeren over de actuele status van de arbeidsmarkt. De *Los Angeles Times* gebruikt hiervoor simpele visualisaties gecombineerd met korte inhoudelijke tekstuele analyses, in de vorm van een *quick take article* (Lee, 2015). Ook Forbes volgt deze lijn, en probeert een kort, maar duidelijk, beeld te geven van de actuele situatie (Sharf, 2015). Een uitgebreidere visuele analyse komt van de *Washington Post*, met hun *Jobs Report* (Cameron & Park, 2015). De meest uitgebreide analyse van de actuele status van de arbeidsmarkt komt van de *Wall Street Journal*, met hun *Job Market Tracker*. Deze visualisatie heeft als doel een goed overzicht te geven van de werkeloosheid op het niveau van specifieke sectoren, en op nationaal niveau.

Dit paper analyseert deze visualisatie, en probeert inzichten te geven in de manier waarop multidimensionale concepten als werkeloosheid en arbeidsmarktperspectieven gestructureerd getoond kunnen worden. Dit paper is als volgt georganiseerd. Allereerst wordt het theoretische kader uitgelegd, waarbij er wordt gekeken naar twee analytische modellen van Edward Tufte, het conceptuele kader (CRAP) van Robin Williams, en het concept van *visual encodings*. Vervolgens wordt de casus van de *Job Market Tracker* getoetst op basis van deze concepten. Afsluitend wordt het succes van de visualisatie bekeken ten

opzichte van het beoogde doel, en worden er suggesties gedaan voor volgende analyses.

## 2 Theoretisch kader

Na een inleiding van het algehele stuk wordt het theoretische kader ontwikkeld. Hiervoor worden er allereerst twee analytische modellen van Edward Tufte geïntroduceerd, de *principles of graphical integrity* en *visualization design principles*. Ook wordt het CRAP-model van Robin Williams uitgelicht, en wordt het concept van *visual encodings* kort uitgelegd.

### 2.1 *Principles of Graphical Integrity*

In zijn boek *The Visual Display of Quantitative Information* noemt Edward Tufte zes principes waar een kwantitatieve visualisatie aan moet voldoen. Allereerst is het belangrijk dat de getallen die gepresenteerd worden proportioneel zijn aan de getallen die gerepresenteerd worden. Ten tweede stelt hij dat assen op de juiste manier gelabeld moeten worden, om ambiguïteit in de data te voorkomen. Het derde onderdeel stelt dat uitleg van de visualisatie ingebed moet worden in de visualisatie zelf, niet ergens anders. Als vierde element noemt Tufte dat variaties aangebracht moeten worden in de data, niet in het design. Het vijfde element omvat de notie dat de visualisatie nooit meer dimensies moet bevatten dan de data. En als laatste element noemt hij dat de data altijd contextueel gebonden moet zijn (1983, p.55-57).

Deze elementen zijn volgens Tufte essentieel voor de integriteit van een visualisatie. Hij noemt hierbij ook een *lie factor*. Deze factor is het resultaat van het delen van de grootte van een effect in de visualisatie door het daadwerkelijke effect in de data. Als deze verhouding hoog is, is er een grote discrepantie tussen de twee elementen, en is de integriteit van de visualisatie laag te noemen.

## 2.2 Visualisation Design Principles

Tufte noemt in ditzelfde boek ook een aantal designprincipes die een designer kan gebruiken om er voor te zorgen dat de visualisatie zogenoemde *chartjunk* bevat, informatie in een visualisatie die niet direct het doel van de visualisatie ondersteunt (1983, p.35). Voor dit paper worden vier van deze elementen gebruikt, namelijk: *micro/macro design layering and separation*, *small multiples* en het principe van *color and information* (1983, p.35). Het principe *micro/macro design* kijkt naar de verhouding tussen kleine elementen (micro) als onderdeel van een grotere structuur (macro), waarbij het belangrijk is om te verklaren door detail toe te voegen (1983, p.37).

*Layering and separation* is een ander belangrijk principe. Het stelt dat verschillende lagen van informatie van elkaar moeten verschillen op basis van textuur, gewicht, vorm, waarde, grootte of kleur (1983, p.58). *Small multiples* behandelt een ander aspect, namelijk de grootte van de informatie die aan gebruikers getoond wordt. Hij raadt hier aan om informatie in kleine, repetitieve onderdelen te presenteren, in plaats van in één keer (1983, p.67).

*Color and information* kijkt naar de manier waarop kleur gekoppeld wordt aan informatie. Tufte noemt hierbij dat designers gebruik moeten maken van lichte versies van blauw, geel en grijs, en dat deze niet sterk contrasteren met de achtergrond (1983, p.90).

## 2.3 Graphic Design Principles

Een andere onderzoeker naar de manier waarop informatie getoond kan worden in visualisaties is Robin Williams. In zijn boek ontwikkelt hij een analytisch model, gebaseerd op vier principes: contrast, herhaling, uitlijning, en afstand. Het eerste principe, contrast, stelt dat elementen die van elkaar verschillen, in een visualisatie sterk verschillend moeten worden getoond (in kleur, dikte, etc). Het principe van herhaling omvat de notie van herkenningspunten. Zorg ervoor dat je gelijke elementen zoveel mogelijk herhaald, om voor een uniforme visualisatie te zorgen.

Uitlijning is een ander belangrijk principe. De locatie van elementen binnen een visualisatie verdient veel aandacht, omdat zo de verschillende elementen *visual connections* met elkaar kunnen maken. Het laatste principe betreft de afstand. Elementen die bij elkaar horen moet bij elkaar worden getoond, en elementen die ver-

schillen, die moeten uit elkaar worden getoond. Dit zorgt ervoor dat de visualisatie niet warrig wordt (Williams, 2003).

## 2.4 Visual encodings

*Visual encodings* omvatten de manier waarop je data is omgezet naar een specifieke dimensie. Hiervoor kunnen verschillende manieren worden gebruikt, zoals positie, grootte, kleur, vorm en textuur. Ieder van deze elementen hebben specifieke eigenschappen. Sommige *encodings*, zoals bijvoorbeeld positie, grootte en lengte, hebben de mogelijkheid om geordend te worden. Sommige *encodings*, zoals kleur en vorm hebben de mogelijkheid om categorische data weer te geven. De keuzes die hierin worden gemaakt zijn belangrijk, omdat ze ervoor zorgen dat de visualisatie goed leesbaar blijft (Iliinsky, 2011).

## 3 Analyse

Nu het theoretische kader is ingeleid, wordt de visualisatie geanalyseerd. Allereerst aan de hand van de twee modellen van Tufte, vervolgens volgens het model van Williams, en uiteindelijk wordt er gekeken naar de manier waarop de informatie wordt omgezet naar het scherm, via *visual encodings*.

### 3.1 Principles of Graphical Integrity

Tijdens de ontwikkeling van de visualisatie is er goed gelet op de proportionaliteit van de data. Het weergeven van groei en afname over de tijd heen van banen in een specifieke sector is gestructureerd gedaan, en is direct proportioneel aan de data die wordt gerepresenteerd. Dit is beide het geval bij de absolute waarden, in het tweede gedeelte van de visualisatie, en bij relatieve waarden, in het eerste deel van de visualisatie. Ook als er hierin gefilterd wordt voor een specifieke groep inwoners, blijft deze eigenschap overeind.

De assen van de visualisatie zijn duidelijk gelabeld, met een korte uitleg in de visualisatie waar nodig (bijvoorbeeld bij het herordenen van de kolommen in het tweede deel van de visualisatie). Over de gehele visualisatie betekent de X-as de groei, absoluut of relatief, over de tijd heen. De Y-as verschilt per onderdeel, maar wordt duidelijk aangegeven.

De visualisatie heeft ook geen uitleg nodig. Er zijn enkele regels uitleg over de specifieke details (zoals over het herordenen), maar voor de rest is

er geen begeleidende tekst. Dit is ook niet nodig om de visualisatie te gebruiken.

Over de variaties binnen de visualisatie is ook goed nagedacht. De enige manier waarop de visualisaties van elkaar verschillen, is de manier waarop de data wordt gepresenteerd. De kleurstelling, tekstuele weergave van elementen, en ordening is over de gehele visualisatie gelijk. Dit zorgt voor veel overzicht.

Het concept dat wordt gepresenteerd is multi-dimensionaal, maar in de gehele visualisatie worden steeds twee dimensies gebruikt: tijd, en de absolute of relatieve groei van banen in de sector.

Dat de data contextueel gebonden gebruikt wordt mag duidelijk zijn. De context van de data is helder over de gehele visualisatie, en door de eenvoud van de data die wordt weergegeven is de verbondenheid met de context de gehele tijd duidelijk.

### 3.2 *Visualisation Design Principles*

Voor deze analyse kijken we naar de vier gekozen elementen van de *visualisation design principles*. Allereerst wordt er gekeken naar de *micro-macro design*. De visualisaties bestaan uit een groot aantal kleine elementen, die samen een groot geheel vormen. Hier hebben de makers van de visualisatie goed over nagedacht. De algehele visualisatie zie er gestructureerd uit (macro-niveau), maar ook op micro-niveau zijn de details duidelijk. Dit wordt bijgestaan door de interacties die plaatsvinden als er voor een specifieke sector wordt gekozen, als voorbeeld. De specifieke details die voor die selectie nodig zijn worden uitgelicht, wat een betere interpretatie van de data als gevolg heeft. Ditzelfde geldt ook voor het derde element van de visualisatie, de *heat map*.

Het principe van *layering and seperation* is ook sterk gedaan. De verschillende lagen informatie worden verdeeld over vier verschillende visualisaties, met vergelijkbare opbouw. Dit zorgt ervoor dat het idee van verscheidenheid tussen de verschillende elementen sterk aanwezig is.

Het principe van *Small Multiples* is in de visualisatie niet sterk aanwezig. De beschikbare data wordt vrijwel altijd in één keer getoond, zonder dat de visualisaties opgebroken zijn in verschillende elementen. Alle sectoren worden over alle jaren heen getoond, waardoor het moeilijk is om een specifieke sector te zien. Het feit dat de eerste twee elementen van de visualisatie aan elkaar gekoppeld zijn (een interactie met het tweede element zorgt voor interactie met het eerste element),

helpt hierin niet mee. Daarvoor is de positionering van de elementen verkeerd (vooral voor gebruikers met een lagere resolutie).

*Color and information* is een sterk element in de visualisatie. De kleuren zijn duidelijk (groen betekent positief, rood betekent negatief), en wordt over de gehele visualisatie op de zelfde manier gebruikt. Dankzij de duidelijke legenda's is er meteen duidelijk welke informatie er schuilt onder een specifieke kleur.

### 3.3 *Graphic Design Principles*

Voor een alternatieve analyse wordt er gekeken naar de vier *graphic design principles* van Robin Williams. Het contrast is in de visualisatie sterk aanwezig. De elementen die niet bij elkaar horen, contrasteren erg op basis van vorm en kleur. Dit is vooral duidelijk bij de 'recessieaanduidingen' in de visualisaties. Ieder element dat niet bij de twee dimensies horen die worden gepresenteerd worden op een geheel andere manier gepresenteerd.

Het herhalingselement is over de hele visualisatie heel sterk aanwezig. Er worden dezelfde kleuren gebruikt (ook bij de interacties), en de twee afzonderlijke groepen visualisaties (op sector-niveau en nationaal-niveau) zijn ook op dezelfde manier aan elkaar gekoppeld.

Uitlijning is wat minder succesvol in deze visualisatie. Als voorbeeld hiervan kunnen de laatste twee elementen gebruikt worden. Hoewel ze dezelfde periode weergeven, zijn de jaartallen niet uitgelijnd. Dit had een positief effect kunnen hebben op de uniformiteit van de algehele visualisatie. Ook begint de visualisatie pagina-breed, maar eindigt relatief nauw. Dit zorgt ervoor dat de focus alleen op het eerste element van de visualisatie ligt, terwijl het laatste element niet minder interessant is.

De locaties van elementen zijn relatief arbitrair. Bij sommige elementen staat de legenda aan de linkerzijde van het element, soms aan de rechterzijde. Hier lijkt ook minder goed over nagedacht te zijn.

### 3.4 *Visual encodings*

Als laatste analyse wordt er kort gekeken naar de *visual encodings* die worden gebruikt in de verschillende elementen van de visualisatie. Als belangrijkste *visual encoding* is er veel kleur gebruikt om absolute of relatieve waarden aan te geven. In de eerste drie elementen van de visualisatie worden dezelfde kleuren gebruikt (met uitzondering

van de blauw kleur om de 'meeste groei') aan te geven. De rood-;groen-codering zorgt ervoor dat de data goed te ordenen is, zelfs zonder de specifieke waardes erbij.

Voor het tweede element van de visualisatie, de omgekeerde *bar chart*, wordt niet alleen kleur gebruikt, maar ook lengte. De lengte van het bar-element geeft precies weer wat de absolute waardes zijn van het aantal banen in die sector. Gecombineerd met de kleuren (twee visuele encodings) zorgt dit voor een groot geheel.

## 4 Conclusie

Met de *Job Market Tracker* heeft de Wall Street Journal een sterke visualisatie neergezet, die het direct duidelijk maakt wat de huidige status is van arbeidsmarktperspectieven op nationaal- en sectorniveau. De ontwikkelaars van de visualisatie hebben goed nagedacht over een groot aantal aspecten die belangrijk zijn bij de ontwikkeling van een analyse. Naast een groot aantal sterke punten heeft de visualisatie ook een aantal slechte punten. De elementen op de pagina zijn niet uitgelijnd, en laten soms dezelfde informatie zien. Vooral de laatste twee elementen laten in principe dezelfde informatie zien, waarbij het bij de *line graph* veel duidelijker getoond wordt. De *heat map* maakt het mogelijk om op maand-niveau informatie te zien. Maar, omdat daar geen informatie is (er zijn geen duidelijke verschillen tussen maanden), zou dat weggelaten kunnen worden.