

# Data Processing

## Week 5

---

04-03-2016

Universiteit van Amsterdam

Minor Programmeren

Groep B1

Julia Jansen

Mats van Beelen

Shannon Bakker

Tom Schoufour

Thom Mekelenkamp

## Problem 1: Design critique group reflection

### Wat is het domein of de context van de visualisatie?

Deze visualisatie gaat over de beoordelingen van films van zowel critici als het grote publiek. De doelgroep van de visualisatie bestaat uit mensen die geïnteresseerd zijn in de filmindustrie, maar geen experts zijn.

### Taken van de visualisatie:

- Het verschil laten zien tussen de mening van het publiek en die van critici voor films die gemaakt zijn tussen 2007 en 2011.
- De correlaties tussen budget en waardering weergeven.
- De correlaties tussen opbrengst en waardering weergeven.
- De correlaties tussen het winnen van prijzen en de waardering weergeven.
- De correlaties tussen opbrengsten in binnen en buitenland en de waardering weergeven.
- De correlaties weergeven tussen aan de ene kant de waarderingen van critici en publiek en aan de andere kant het succes en de verdiensten van de film en het aantal bioscopen waar hij werd gedraaid in het eerste weekend.

### Integriteitprincipes:

- De y-as heeft een logische schaal maar is onduidelijk doordat hij niet goed gelabeld is. Je moet de tekst onder de grafiek lezen om de begrijpen dat het om de beoordeling van de films gaat.
- De x-as heeft geen label en er is ook geen schaal weergegeven. Hierdoor moet je als kijker erg lang naar de visualisatie kijken om een patroon te herkennen. Dan wordt langzaam duidelijk dat de x-as van links naar rechts het verschil tussen de beoordeling van de kijkers en die van de critici weergeeft. Links in de grafiek zijn de critici positiever zijn dan het publiek, rechts is het publiek positiever dan de critici. De witte trechter geeft aan hoe groot het verschil is tussen de critici en het publiek voor de films op dat punt
- De *lie factor* is heel laag, maar bij *profitability* is er één film die heel veel heeft opgebracht in verhouding tot het budget. Daardoor komt de schaling zo uit dat de rest van de bolletjes van vrijwel dezelfde grootte zijn, terwijl dat niet overeenkomt met de data.
- De variatie die te zien is in de grafiek is bij de meeste variabelen een weerspiegeling van de data, behalve wanneer het gaat om de winst. Dan is er geen variatie zichtbaar terwijl er wel variatie in de data is. Dit komt doordat er één film is die enorm veel winst heeft gemaakt ten opzichte van het budget (namelijk *paranormal activity 11420X* eigen budget aan opbrengsten). Hierdoor is de variatie tussen de andere films onzichtbaar.

### Designprincipes

- De *data-ink ratio* is goed totdat je op een film geklikt hebt en de achtergrond verandert in de afbeelding van de film. Hierdoor wordt er inkt verspild aan het weergeven van de foto. Gelukkig kan deze functie worden uitgeschakeld.
- De *data density* is goed, want er wordt veel data weergegeven in de visualisatie en er is weinig lege ruimte.
- Interactieve menu's geven de gebruiker de mogelijkheid om acht extra lagen te bekijken. Men kan alleen niet meerdere van die extra lagen tegelijkertijd bekijken.

## Grafische principes

- Er wordt gebruik gemaakt van verschillende kleuren, maar de grafiek zou beter zijn wanneer het contrast tussen de verschillende kleuren hoger zou zijn. Dit wil zeggen dat lichtrode en lichtblauwe bolletjes bijvoorbeeld de voorkeur hebben boven de nu gebruikte roze en paarse bolletjes. Het paars op het zwart is ook niet gelukkig gekozen, omdat paars een kleine golflengte heeft en menselijke ogen hier niet zo gevoelig voor zijn (Ware, H4: pg 67). Een kleur met een wat langere golflengte komt beter uit de verf op een zwarte achtergrond.
- *Alignment* is goed in deze grafiek, want alle data wordt gevisualiseerd op basis van de zelfde assen.
- De *proximity* in de visualisatie is goed omdat het verschil tussen de waarderingen ook bepaald waar een film komt te staan. De films die qua waardering van publiek versus critici op elkaar lijken staat dus dicht bij elkaar in de visualisatie.

## Visual encodings

- Ze hebben bolletjes voor de films gebruikt. Dat is gepast, maar de bolletjes hadden kleiner gekund. Hierdoor zou de grafiek beter leesbaar worden, omdat de bolletjes dan minder zouden overlappen en de positie van de bolletjes ten opzichte van de assen helderder wordt.

## Subjectieve dimensies

- Esthetisch vinden we de grafiek mooi alleen past de witte trechter niet goed bij de rest van de visualisatie.

## Doel

- Inzicht geven in de verschillen tussen de beoordelingen van het publiek tegenover die van de critici. Dat kan vergeleken worden met een aantal variabelen waaronder de opbrengsten van de films en het succes van de films in het openingsweekend.
- Het doel wordt niet optimaal bereikt, want je ziet wel de positie van alle stippen (films) ten opzichte van hun waardering door publiek en critici, maar je ziet steeds van maar één film de titel. Directe vergelijking tussen films wordt daardoor lastig, want de kijker moet telkens de positie van een film onthouden om die met een andere specifieke film te vergelijken.

## Verbeterpunten

- De schalen aanpassen en goed labelen.
- De bolletjes kleiner maken en meer ruimte tussen de bolletjes onderling laten.
- Ook gebruik maken van andere grafiektypen voor verschillende data. Bijvoorbeeld d.m.v. kleine staafdiagrammen onderaan de huidige visualisatie, die de top x aantal films weergeeft voor een bepaalde variabele zouden de films beter direct met elkaar vergeleken kunnen worden. Dan is ook duidelijker te zien welke films op bepaalde variabelen eruit springen.
- Een functie toevoegen die meerdere films met elkaar kan vergelijken.
- Andere kleuren gebruiken, die meer contrast met elkaar en meer contrast met de achtergrond hebben.
- Waardes weergeven van geselecteerde attributen. Nu is niet te zien wat de grootte van een bolletje voor waarde van een bepaalde attribuut weergeeft.

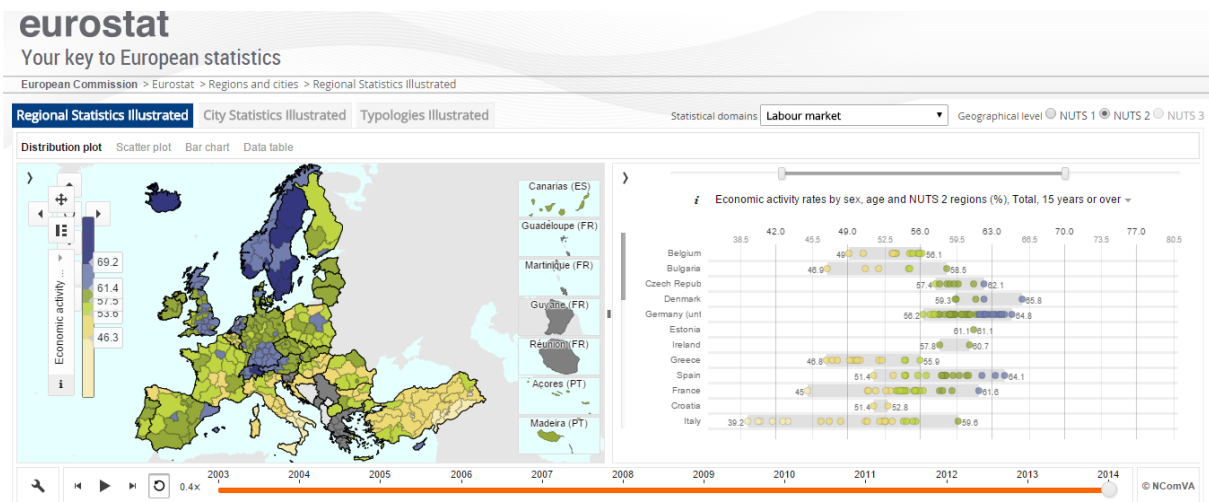
## Problem 2: Patterns and colors in maps

Wij vergelijken de kaarten van Bing en Google.

- Bij Bing Maps is het makkelijker om gebouwen te vinden. Er is een hoog contrast tussen de gebouwen en de omgeving. En wanneer je met je muis over een gebouw van Harvard University hoovert zie je alle gebouwen van de campus.
- Uitgezoomd is de kaart van Google overzichtelijker dan die van Bing, omdat de plaatsnamen pas zichtbaar worden als je verder inzoomt. De kaart is zonder die vele plaatsnamen rustiger voor het oog, waardoor de aangegeven route duidelijker is op de kaart. Daarbij heeft Google meer informatie over mogelijke routes met verschillende soorten vervoer zoals de fiets en het vliegtuig.
- We vinden de kaarten van Google prettiger bekijken dan die van Bing. Er is meer interactie mogelijk en er is meer informatie te vinden op de kaart. Ook is het contrast tussen verschillende onderdelen op de kaart over het algemeen beter dan het contrast bij Bing Maps. Zo is het oranje van de grote wegen bij Google Maps beter van de lichtgrijze achtergrond te onderscheiden dan de lichtpaarse wegen bij Bing Maps. Over het algemeen is de visuele weergave van wegen, bebouwing, groen en water bij Google Maps helderder en zijn de visuele attributen van betere kwaliteit. De verschillende onderdelen bij Bing Maps lijden een beetje onder pixelige kwaliteit.

## Problem 3: Multiple Linked Views Visualisation

<http://ec.europa.eu/eurostat/cache/RSI/#?vis=nuts2.labourmarket&lang=en>



### Publiek

Het doel van deze visualisatie is inzicht geven in een scala aan data (hierboven genoemd) over regio's binnen Europa. Eurostat wordt voornamelijk gebruikt door beleidsmakers en journalisten, maar het moet ook begrijpelijk zijn voor alle Europese burgers. Het publiek moet wel naar een specifiek onderdeel van alle data op zoek zijn om echt profijt te hebben van deze multiple views visualisatie.

## Onderdelen/doelen

- De kaart geeft per regio in Europa aan wat te waarde is voor een bepaalde variabele die de gebruiker heeft gekozen in het rolmenu boven de rechterafbeelding. In de kaart worden de waardes met verschillende kleuren weergegeven. De kleuren zijn goed gekozen, want ze lopen van licht naar donkerder en van kleuren met een hoge golflengte naar kleuren met een kleine golflengte, op een lichte achtergrond. Dit brengt in juiste volgorde de hoeveelheid van een variabele over op de lezer (Ware, H4: pg 84).

- Het distribution plot ernaast geeft bijvoorbeeld per land aan hoe groot het percentage werkzame mensen boven 15 jaar in de regio's van dat land is. Door middel van een grijze balk wordt de range per land aangegeven. De gebruiker kan boven de landkaart aanklikken wat hij/zij rechts wil zien: een distribution plot, een scatterplot, bar chart of data table. De linker afbeelding blijft telkens een geografische weergave, maar de kleuren veranderen al naar gelang welke variabele de kijker in geïnteresseerd is.

- Er valt in de rechter grafiek bovendien ook nog te kiezen welk deel van de bevolking men in geïnteresseerd is. Als de gebruiker op regio's in de linker afbeelding klikt, verschijnt een tekstblokje bij de muis waarin de data wordt weergegeven die in de rechterafbeelding is gekozen. Dit correspondeert goed met elkaar, beide afbeeldingen veranderen direct mee als er iets door de gebruiker in de ene of andere afbeelding wordt veranderd.

- Als je met je muis klikt op een regio krijg je de data van die specifieke regio te zien. Ook wordt de regio op de plot ernaast omcirkeld. Hierdoor weet de gebruiker welke data uit de plot bij de geselecteerde regio hoort. En dat werkt ook andersom. Dit type interactie komt volgens Baldonado, Woodruff en Allan Kuchinsky (2000) regelmatig voor in multiple linked views visualisations . Zij noemen deze techniek 'brushing'.

- Wanneer scatterplot en de geografische kaart naast elkaar te zien zijn, geeft dit de kijker daadwerkelijk meer informatie dan wanneer de twee afbeeldingen los van elkaar of na elkaar te zien zouden zijn. Als er op het pijltje bovenaan de plot wordt geklikt kan er een reeks visuele instellingen worden veranderd aan de plot. De veelheid aan attributen (geografische ligging, economische activiteit, werkloosheid, bevolkingsdichtheid, toerisme, gezondheid, onderwijs, transport, agricultuur, demografie, internetgebruik en wetenschappelijke ontwikkeling) en veelheid aan data (leeftijden, dichtheden, geldbedragen, bezit van goederen en nog veel meer) kan door deze visualisatie beter worden onderzocht door de kijker, dan middels een single view visualisatie. Het is waardevol om rechts te zien hoe de graad van een gemeten variabele geografisch verdeeld is en dan links te kunnen zien waar die regio zich ten opzichte van alle andere regio's bevindt op basis van desbetreffende variabele. Die connectie stelt de lezer in staat de data beter te begrijpen en verbanden te leggen die anders niet mogelijk waren.

- Er is een tijdlijn aanwezig onderaan de visualisatie. Hier kan de lezer zelf een tijdstip aangeven waarop hij wilt kijken. Als er op de play-knop wordt gedrukt zal er een animatie beginnen die vanaf geselecteerde punt op de tijdlijn tot aan het laatste moment op de tijdlijn op de kaart de data weergeeft. Deze tijdlijn kan ook handmatig verschoven worden. Het is door dit tijdsaspect van de visualisatie nog beter duidelijk hoe de data in de twee afbeeldingen met elkaar samenhangt omdat gebruikt wordt gemaakt van beweging en kleur (. Zowel de geografische gebieden kleuren naarmate de kenmerken van die regio veranderen, als ook de plaats van de landen/bolletjes in de scatterplots of andere grafieken die rechts worden weergegeven.

- Je kan de x-as van de plot aanpassen om de percentages die zichtbaar zijn te veranderen.

## **Bibliografie**

Colin Ware (2008) 'Chapter 4, Color' in Visual Thinking for Design. Pp. 65 – 85.

Michelle Q. Wang Baldonado, Allison Woodruff & Allan Kuchinsky, "Guidelines for Using Multiple Views in Information Visualization", Proceedings of Advanced Visual Interfaces (AVI 2000), pages 110-119. ACM Press, 2000.