

## Data Processing

### Design 3

Bob Keijzers

#### Problem 1: Design critique

Het eerste dat mij opvalt is dat de visualisatie er erg rommelig uit en het schema is vol met verschillende soorten felgekleurde bolletje en dit zorgt voor het drukke aanzicht.

Het onderwerp van de visualisatie is hetzelfde als in de 'Eb en vloed' visualisatie van de NY Times, namelijk films. In dit geval gaat het om een verhouding tussen het budget en de bruto omzet van individuele films. De dataset bestaat uit 600 Hollywood films van het afgelopen vijf jaar met de financiële gegevens en beoordelingen van zowel critici als het reguliere publiek. De y-as symboliseert de hoogte van de beoordeling van het publiek (roze) en de critici (purple). De X-as geeft het verschil tussen critici beoordeling en publieksbeoordeling, rechts een positief verschil, links een negatief verschil.

Er kunnen diverse taken worden uitgevoerd binnen deze visualisatie. In het rechter menu wordt de optie 'resize movies by' gegeven, waarbij op meerdere variables kan worden geselecteerd. Zo kunnen de publieks- en critici beoordelingen van de films worden bekeken, deze zijn geschaald op de y-as. Winstgevendheid, budget, omzet in het binnen- en buitenland en wereldwijd, het aantal bezoekers in de eerste week en de winst en het succes in de eerste week. Ook kan er geselecteerd worden om alleen films weer te geven waarin critici en publiek het qua beoordeling eens zijn of alleen films met hetzelfde beoordeling.

Het menu waarin films met dezelfde rating kunnen worden weergegeven is niet correct gelabeld, de betekenis van het eerste symbool is niet geheel duidelijk, de overige symbool zijn wel correct gerangschikt. De categorieën van de beoordelingen zijn correct gerangschikt.

Het valt gelijk op dat er moeilijk is om over de verschillende bolletjes te 'hooveren', het is niet eenvoudig de muis op specifieke bolletje te houden zodat er informatie over de film verschijnt.

Het grootte van de bolletjes wanneer binnenlands, buitenlands en wereldwijde omzet wordt bekeken is echt klein. De grootte van de bolletjes bij selectie op aantal toeschouwers in de eerste week is te groot, waardoor er niet correct kan worden afgelezen. De grootte van de bolletjes is incorrect wanneer winstgevendheid en succes in de eerste week wordt geselecteerd.

De data-ink ratio is relatief hoog, wanneer de begeleidende tekst niet wordt meegerekend. De information dichtheid is hoog. Al hoewel beoordelingen en financiële gegevens gezamenlijk worden weergegeven kan er goed gefilterd worden op diverse variables. Contrast is hoofdzakelijk aanwezig door de grootte van de bolletjes.

De visuele codering is niet heel uitgebreid, er wordt slechts gebruik gemaakt van twee kleuren en een vlak dat het gemiddelde verschil tussen critici en publiek beoordeling aangeeft.

Alhoewel de data visualisatie er op het eerste oog rommelig uitziet, geeft deze eigenlijk erg veel informatie en wordt het doel van de visualisatie bereikt.

## Problem 2: Rainbow color map

Onderstaande regenboog visualisatie is gebruikt voor deze individuele opdracht. Deze grafiek geeft het percentage van neerslag weer dat wordt opgenomen voor planten in de Verenigde Staten in periode 1971-2000.

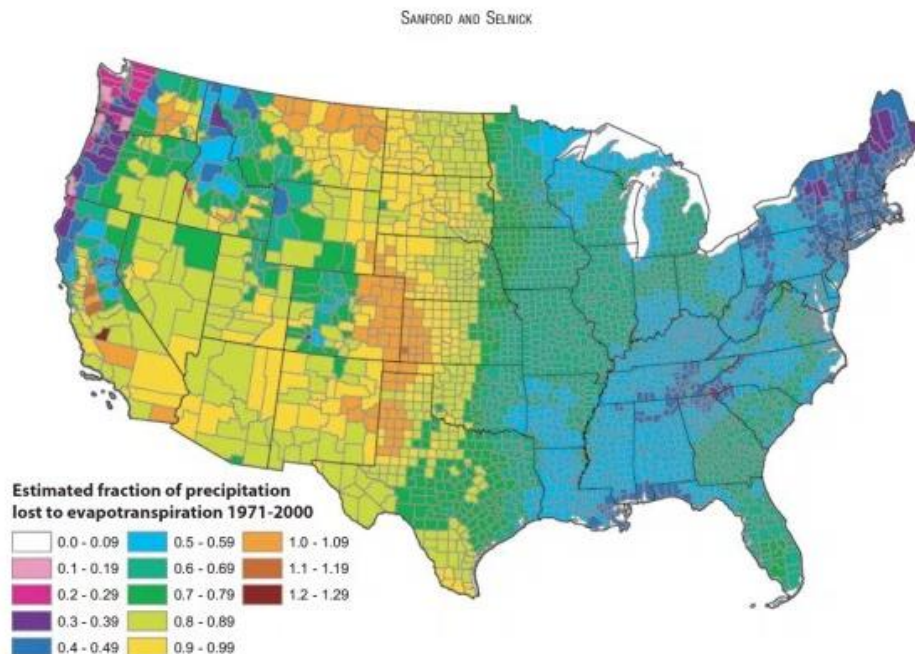


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

LINK: <https://eaqereyes.org/basics/rainbow-color-map>

Deze manier van visualiseren doet mij gelijk denken aan atlassen, om de verschillende tussen de regio's van bepaalde variables weer te geven, wordt de rainbow visualisatie dikwijls ingezet.

Dit is een goede voorbeeld van een misleidende rainbow visualisatie. Op het eerste gezicht lijkt er een groot verschil te zijn tussen het oosten en westen van de Verenigde staten. Dit geeft gelijk een duidelijk beeld. Echter wanneer de legenda goed wordt bekeken, blijken de dominante geel en lichtgroen tint (het westen) en de dominante groen en lichtblauw tint (het oosten) qua waarde heel dicht bij elkaar te liggen. Het verschillen zijn dus helemaal niet zo groot als op het eerste gezicht wordt gedacht.

Hieruit kan worden geconcludeerd uit de overloop van de verschillende kleuren niet erg goed is, als de verschillen tussen de aan elkaar grenzende tinten groot zijn, worden de werkelijke verschillen tussen de regio's overschat. Tevens kan er niet intuïtief gezien worden waar de waardes hoog en laag zijn, dit kan bij veel andere visualisaties wel en is daarom een tekortkoming van de rainbow visualisatie.

De visualisatie geeft objectief gezien geen verkeerd beeld, want de juiste data kan weldegelijk worden afgelezen in de visualisatie. Er is geen duidelijk reden waarom er juist voor een rainbow color scale is gekozen. Echter, zoals hierboven besproken, geeft de grafiek een misleidend beeld waarin de

verschillen tussen regio's worden overschat. Een geschikte visualisatie is wat mij betreft een visualisatie met een scala van tinten van licht naar donker, van dezelfde kleur. Hierin visualiseert lichtblauw een lage score en donkerblauw een hoge score.