Preliminary portfolio 05/10/15 Peter van Twuijver en Sophie Hutten

We hebben het hele stappenplan van ons project uitgestippeld en in een script geplaatst, maar merken dat we tussendoor veel aanpassingen hebben moeten doen. We zijn tegen een aantal dingen aangelopen die we aan hebben moeten passen.

* Toen we probeerden om de postcodes in R.Studio te ‘runnen’ kregen we een foutmelding:   
  **“Error: google restricts requests to 2500 requests a day for non-business use”**
* Dit betekent eigenlijk dat we het sowieso op moeten delen in verschillende databestanden, waar we niet op hadden gerekend.
* We hadden de landen (NL en BE) weggelaten, maar zijn er achter gekomen na wat experimenteren met R Studio dat de landen nodig zijn om de postcode juist te vinden. (bijv. Postcode 1019 wordt anders aangeduidt in Phang Phieng). Deze hebben we bij de stap waar we de letters van de postcode verwijderen ook weggelaten. Nu proberen we uit te vinden hoe we ervoor kunnen zorgen dat R Studio alleen in Nederland zoekt naar de aangegeven postcodes (Als we achter de postcodes ‘Nederland’ schrijven doet de postcode het namelijk wel). Om dit te testen hebben we een testbestand aangemaakt (testje.txt). In dit bestand hebben we alleen rij 1 tot en met 200 gebruikt, we hebben dit geknipt door middel van command line.

We zijn redelijk klaar met het gedeelte dat we via terminal hebben gedaan, het aanpassen van de data. Nu wilden we aan de slag in R maar liepen we dus tegen deze uitdaging aan. We gaan nu nadenken hoe we onze data willen ‘knippen’ in verschillende bestanden, omdat het anders door Google niet kan worden gelezen. Ook willen we een oplossing vinden voor het probleem dat sommige postcodes niet binnen Nederland worden gevonden. Als we dit hebben opgelost kunnen we beginnen met de data daadwerkelijk omzetten in locaties en hier een map van maken. Het plan is om een tijdsgebonden visual te maken van de verschillende postcodes en de bijbehorende data. Als het niet lukt, moeten we de postcodes wellicht wel met letters erachter doen… alleen zijn we dan bang dat we te veel kleine ‘puntjes’ krijgen.   
  
Wat we tot nu toe hebben geleerd over de onderzoeksvraag is eigenlijk nog niets. We zijn nu vooral bezig geweest met het aanpassen van de data en nu met het verwerken van de data. Maar zodra dat is gelukt, denken we een stuk wijzer te zijn wat betreft de onderzoeksvraag: “Waar en wanneer was de hype het grootst, en hoe is de populariteit van de selfie-stick vandaag de dag?”   
Zodra we het visueel hebben, hopen we antwoord te kunnen geven op de volgende vragen:  
- Kunnen we duidelijke pieken aangeven wat betreft verkoop? (Wellicht kerst/vakantie?)  
- Is er een duidelijk verschil te zien binnen bepaalde gebieden?