



Cursus: Data Visualisation & Analysis (Hogeschool Utrecht)
Student: Adriaan Stolk
Studentnummer: 1517495

Inhoudsopgave

Inleiding.....	3
Datasets	3
Conclusies:	4
Visualisatie 1: Comparing categories	5
Visualisatie 2: Timeseries	6
Visualisatie 3: Distribution	7
Visualisatie 4: Relationships	8
Visualisatie 5: Clustering	10

Inleiding

Voor de cursus Data Visualisation & Analysis aan de Hogeschool Utrecht is er een analyse gemaakt van twee WhatsApp-gesprekken. De ene is een groepschat van een hockeyteam en de ander is een één-op-één gesprek met een partner.

Datasets

1. Groepschat hockeyteam

Dataset van een groepschat van het hockeyteam van Voordaan Heren 1. Dit is een hockeyteam op het één-na-hoogste niveau in Nederland, wat af en toe een uitstapje maakt naar de Hoofdklasse (hoogste niveau). Aan het eind van het seizoen zijn er vaak wisselingen in de groepschat, omdat er spelers stoppen en nieuwe spelers bijkomen.

Aantal berichten:	11.442
Aantal deelnemers:	72
Periode:	17-05-2020 tot en met 18-02-2025
Gebruikt voor:	Comparing categories, relationships en clustering

De data is verrijkt met een Roles.json-file. Hierin wordt voor elke auteur van de groepschat aangegeven welke functie de auteur heeft binnen de groep. Dit kan zijn speler of staff. Onder staff verstaan we eigenlijk alle personen die betrokken zijn bij het team, die geen speler zijn. Dit zijn bijvoorbeeld de fysio, de coach, de assistentcoaches, de manager, de videoman of de looptrainer.

Daarnaast wordt dit ook verrijkt met de positie van een speler binnen het veld. (bijvoorbeeld: een keeper of een verdediger). Als iemand onderdeel is van de staff, is dit de positie.

2. Chatgesprek met vrouw

Dataset van een één-op-één gesprek met mijn vrouw.

Aantal berichten:	99.998
Aantal deelnemers:	2
Periode:	18-08-2020 tot en met 11-02-2025
Gebruikt voor:	Time series en distribution

Conclusies:

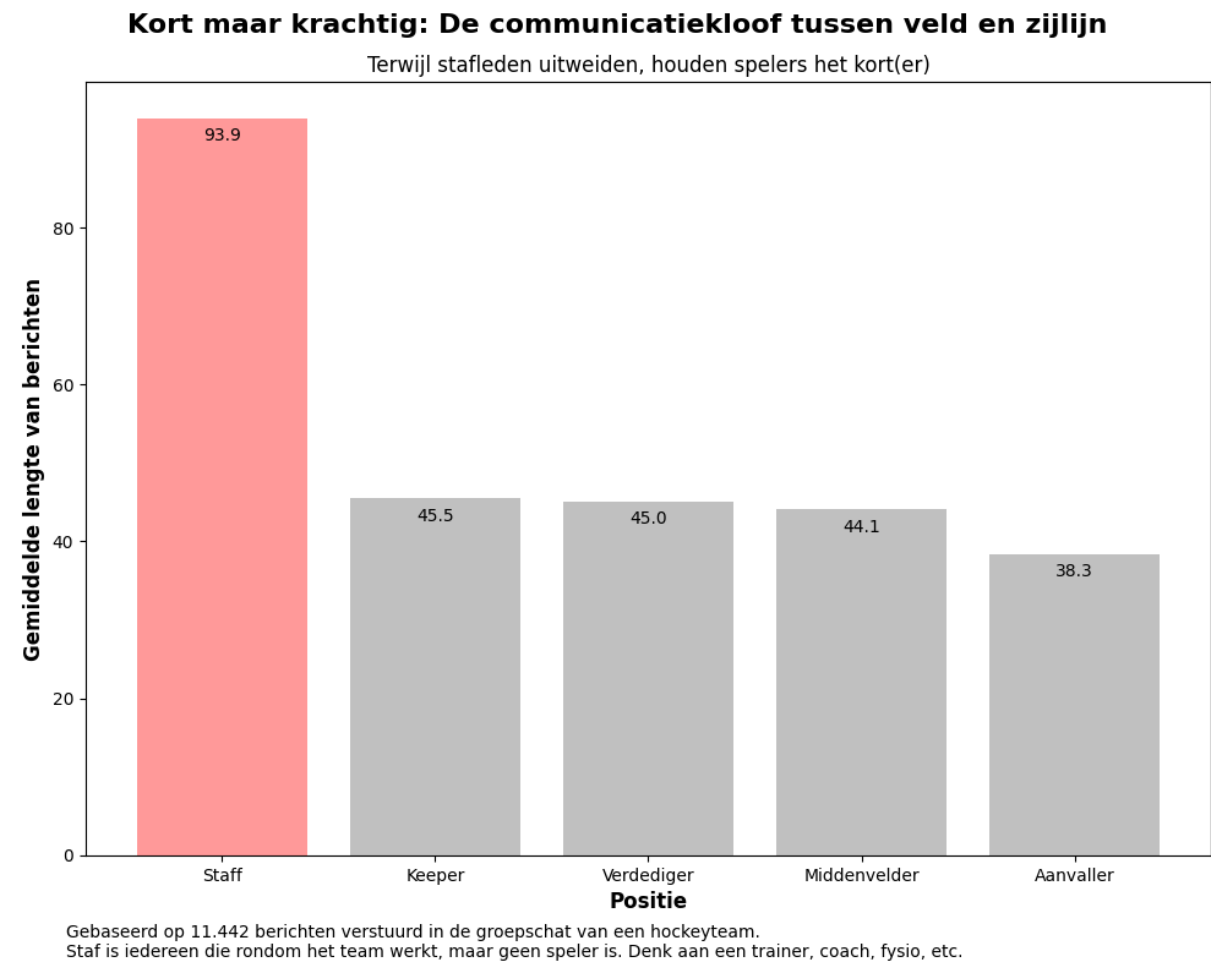
Er zijn een aantal conclusies die getrokken kunnen worden op basis van hetgeen we in de visualisaties zien. Zo komt duidelijk naar voren in de visualisaties van het hockeyteam dat er overduidelijk verschil zit in de manier van communiceren van stafleden ten opzichte van spelers. Dit gaat dan niet alleen om de lengte van berichten (zie comparing categories visualisatie), maar ook over de inhoud van de berichten van met name de hoofdcoach (zie clustering visualisatie).

Maar ook binnen de spelersgroep zelf is er duidelijk een verschil in activiteit in de groepschat. Het valt op dat de verdedigers het meeste reageren en dat stafleden met name op elkaar reageren (zie relationships visualisatie).

Bij de chats met mijn vrouw is de conclusie overduidelijk dat we ontzettend snel op elkaar reageren. 73,4% (!) van de berichten worden beantwoord binnen een kwartier. Ook is het duidelijk te zien dat we in de aanloop naar en na de geboorte van onze dochter percentueel gezien meer foto's zijn gaan sturen.

Al deze conclusies zijn niet zichtbaar als je onderdeel bent van een chat. Sommige conclusies zijn niet zo heel verrassend (een coach die lang van stof is), maar de visualisaties laten zeker verbanden zien die eerder niet opgevallen zouden zijn.

Visualisatie 1: Comparing categories



Afbeelding 1: comparing categories

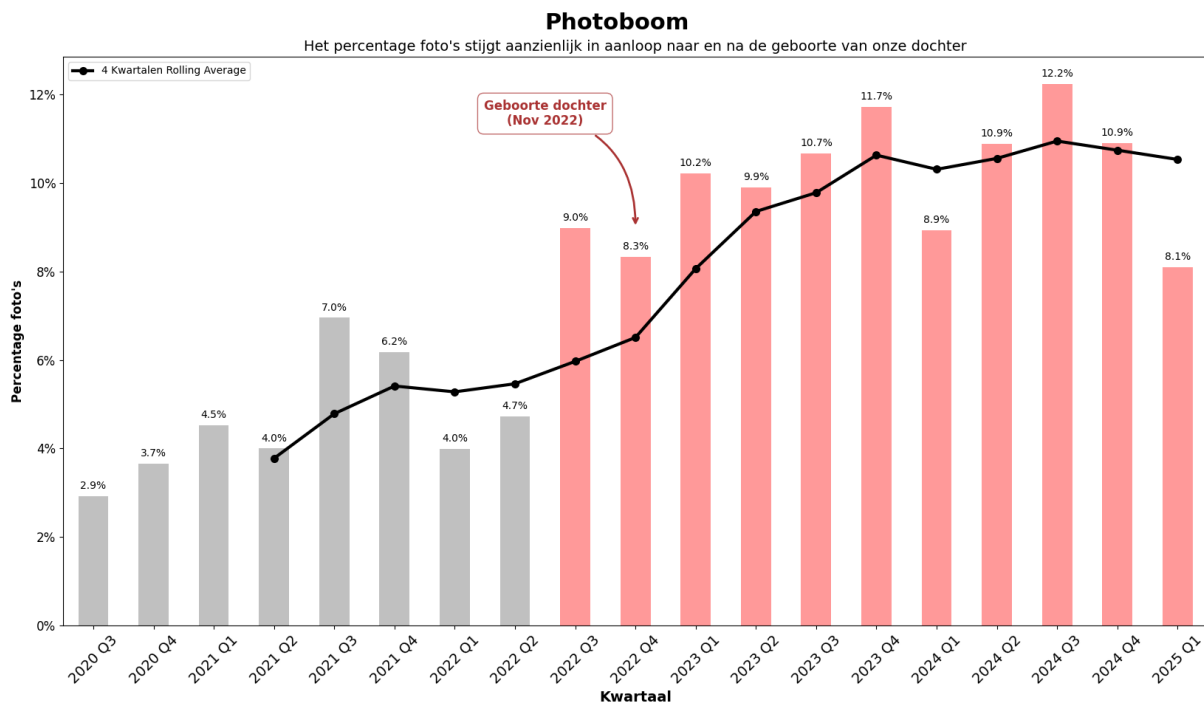
De visualisatie laat in één oogopslag zien wat iedereen in de kleedkamer en op het veld weet: de spelers, vaak jonger en meer met smartphone bezig dan stafleden, communiceren niet alleen anders maar zeker ook korter. Ze hebben beduidend minder tekst nodig om een boodschap naar elkaar over te brengen. Een andere verklaring is dat de staf meer berichten stuurt over verzamelen, voorbereiding, wedstrijden en dat spelers meer bezig zijn in het uitvoeren van hun primaire taak: hockeyen.

Gehanteerde principes:

De bar van de staff is in de “huisstijl” kleur opgemaakt, zodat hij meteen eruit springt ten opzichte van de grijze balken (preattentive processing). Eerder had ik de grafiek gemaakt met alleen de uitzetting op basis van functie (dus speler of staff), maar om het verhaal meer kracht te geven en duidelijk te maken dat eigenlijk de spelers op verschillende posities ongeveer dezelfde gemiddelde lengte aan berichten versturen, heb ik ervoor gekozen om dit toch op basis van positie te doen. Dit maakt in mijn ogen het verhaal van de visual sterker.

De begeleidende tekst geeft duiding over de hoeveelheid berichten die verstuurd zijn en wat nou het onderscheid is tussen speler (of positie) en stafleden.

Visualisatie 2: Timeseries



Gebaseerd op 99.998 berichten tussen twee partners tussen 18-08-2020 en 11-02-2025

Afbeelding 2: time series

Ons gezin bestaat uit 4 personen. Mijn partner, haar zoon (mijn “bonuszoon”) en sinds november 2022 hebben mijn partner en ik een dochter samen. Ik had altijd het vermoeden dat wij veel meer foto's naar elkaar gingen versturen een tijdje voor en na de geboorte van onze dochter, dus het idee van deze visual was snel geboren (pun intended).

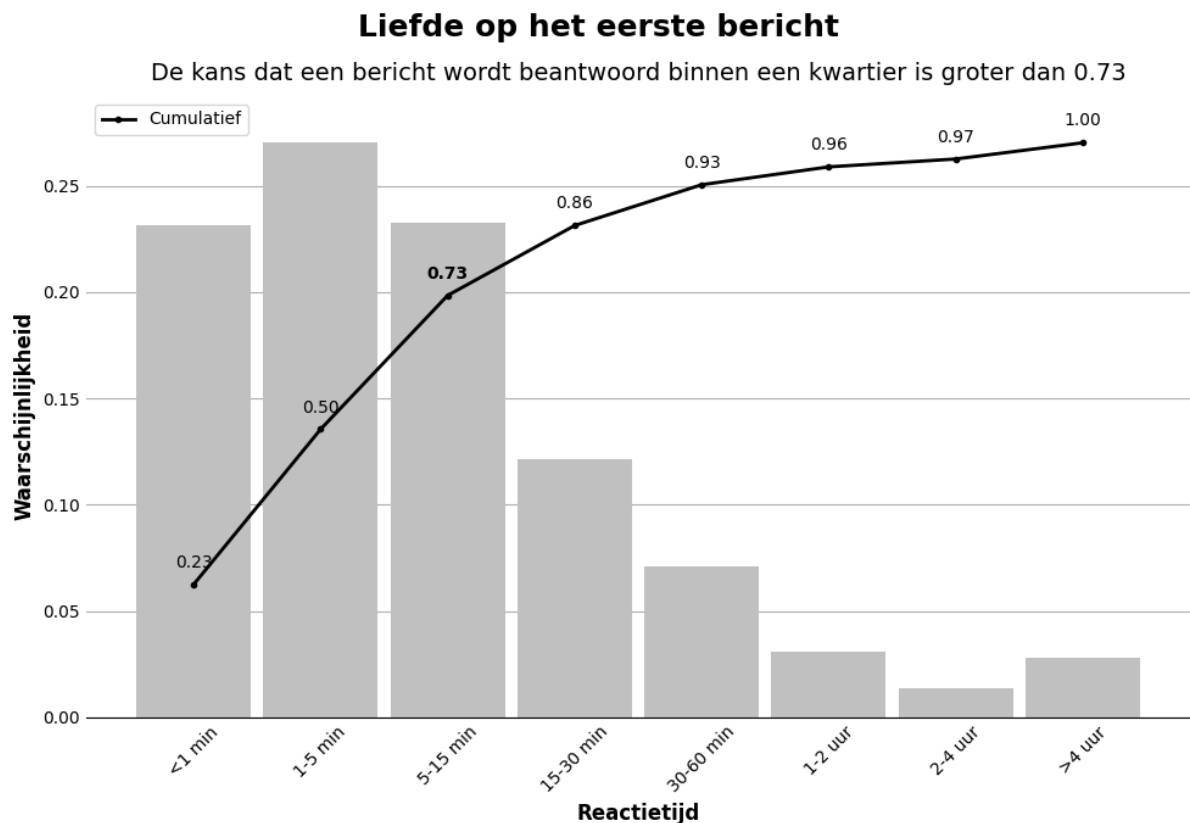
Gehanteerde principes:

Met de kleuren (huisstijl) geef ik meteen de kwartalen van en na de geboorte van onze dochter aan, waaruit duidelijk naar voren komt dat dit ruimschoots hoger is dan de kwartalen daarvoor. Om dit extra kracht bij te zetten is er een rolling average van 4 kwartalen toegevoegd.

De ondertitel maakt duidelijk dat het ook gaat om de periode vóór de geboorte, waarin er (uiteindelijk) meer foto's van babykamers, kledingsetjes en andere aanverwante babyartikelen worden verzonden.

De begeleidende tekst geeft aan over de hoeveelheid berichten die verstuurd zijn, waaruit blijkt dat het niet gaat om een handvol berichten, maar om bijna 100.000.

Visualisatie 3: Distribution



Gebaseerd op een gesprek tussen twee partners.
Totaal aantal berichten met respons van de andere partner: 36.150. Dit betekent geen reactie op jezelf.

Afbeelding 3: Distributie

Dat mijn partner en ik veel appen en snel reageren, dat kon ik wel bedenken, maar ik wist niet dat we de kans op een reactie binnen een kwartier 0.734 zou zijn. Hier was ik zelf ook vrij verbaasd over.

Hierbij is het ook grappig om te zien dat van de 99.998 berichten in de vorige visual er nog maar 36.150 over zijn met reactie van de ander. Er worden dus veel berichten na elkaar verzonden door dezelfde auteur.

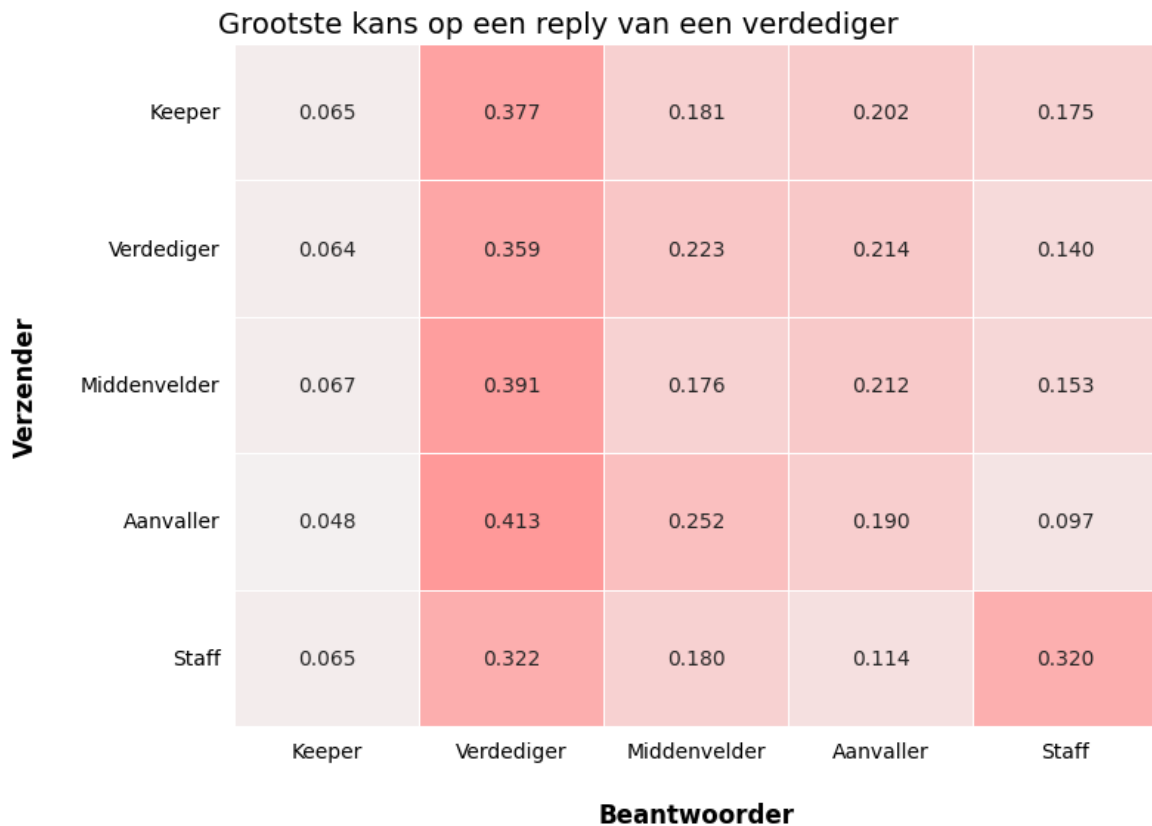
Gehanteerde principes:

Bij deze visualisatie is ervoor gekozen om met name aandacht te besteden aan de lijn, omdat de titel je ook meteen meeneemt naar de 0.73 die bij 5-15 minuten staat. Daarom is het datalabel bij 0.73 ook dikgedrukt.

De tekst onderin de visualisatie legt uit dat het gaat om antwoorden op elkaar, dus niet antwoorden op jezelf.

Visualisatie 4: Relationships

Ook in de groepschat zetten de verdedigers de lijnen uit



Gebaseerd op 9.182 berichten, waarbij iemand niet op zichzelf heeft gereageerd.

Afbeelding 4: relationships

Dit is iets wat je nooit uit een groepschat haalt zonder naar de data erachter te kijken. Iedereen weet dat sommige mensen meer reageren, maar dat het op basis van positie zo duidelijk zou worden, dat had ik niet verwacht.

De oorzaak hierin zou kunnen liggen in het feit dat verdedigers bij het hockey vaak de lijnen uit zetten. Vaak zijn het ook wat oudere spelers, die meer communiceren. Ook zijn twee van de drie (oud-)aanvoerders van het team verdedigers.

De waardes van de keepers zijn natuurlijk vertekend. Toch is er niet één keeper, maar is er vaak ook nog twee reservekeepers. Er zaten in de gehele looptijd 7 keepers in de chat. Dit is wel de minst vertegenwoordigde positie (uiteraard).

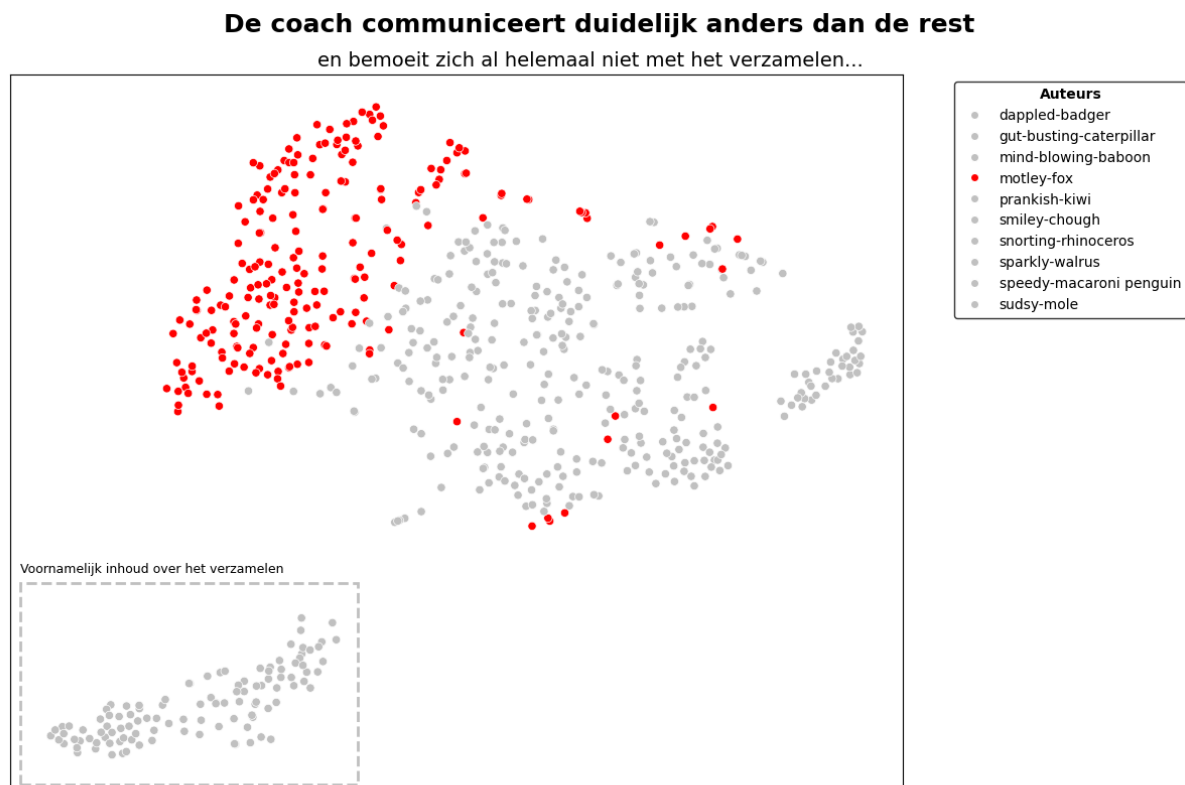
Gebruikte principes:

Op basis van de eerder gebruikte kleur in de andere visualisaties heb ik aan Claude gevraagd om een kleurenpalette te genereren wat gebruikt maakt van aangrenzende kleuren. Ik denk dat het kleurgebruik meteen erg duidelijk laat zien dat de verdedigers het meeste op mensen reageren en dat de staf voornamelijk op elkaar reageert.

De tekst onder de afbeelding illustreert meteen ook om hoeveel berichten het gaat en

dat chats als reactie op jezelf dus niet worden meegenomen in de telling (zie ook het veel kleinere verschil t.o.v. de een-op-een chat).

Visualisatie 5: Clustering



Gebaseerd op de top 10 bijdragers aan de WhatsApp groepchat (op basis van aantal berichten).
Bij het verzamelen wordt meestal één bericht gemaakt, waar de spelers hun eigen naam in zetten en versturen.

Afbeelding 5: Clustering

Motley-fox, onze hoofdcoach, komt in de visualisatie van de clustering mooi naar voren. Hij heeft een hele andere manier van communiceren dan de rest. Dat stafleden anders communiceren dan spelers hebben we natuurlijk al in de andere visualisaties gezien, maar op moment dat je in zo'n WhatsApp groep zit, heb je niet door dat iemand ook zo duidelijk een eigen stijl van communiceren heeft.

Linksonder in de box zie je overigens de berichten die worden gedeeld over het verzamelen. Dan stuurt degene die de verzamelplek regelt een appje door, waarop de rest dit bericht kopieert en zijn of haar naam in plakt.

Gebruikte principes:

Meteen de coach rood eruit laten springen. Dit heb ik eerst geprobeerd te doen met de "huisstijl" rood/roze, maar hierdoor sprong hij er voor mij niet meteen uit als cluster. Met deze kleurstelling wel.

De tekst bij het vlak en de tekst onderaan de pagina verduidelijken wat er voornamelijk in dat cluster zit en geeft ook aan dat dit enkel de top 10 van bijdragers bevat. Eerder kregen we namelijk een lijst van de 72 auteurs en dit zorgde voor erg veel clutter in de visualisatie.