Road sensors and beyond: new data sources in the era of big data

Marco Puts

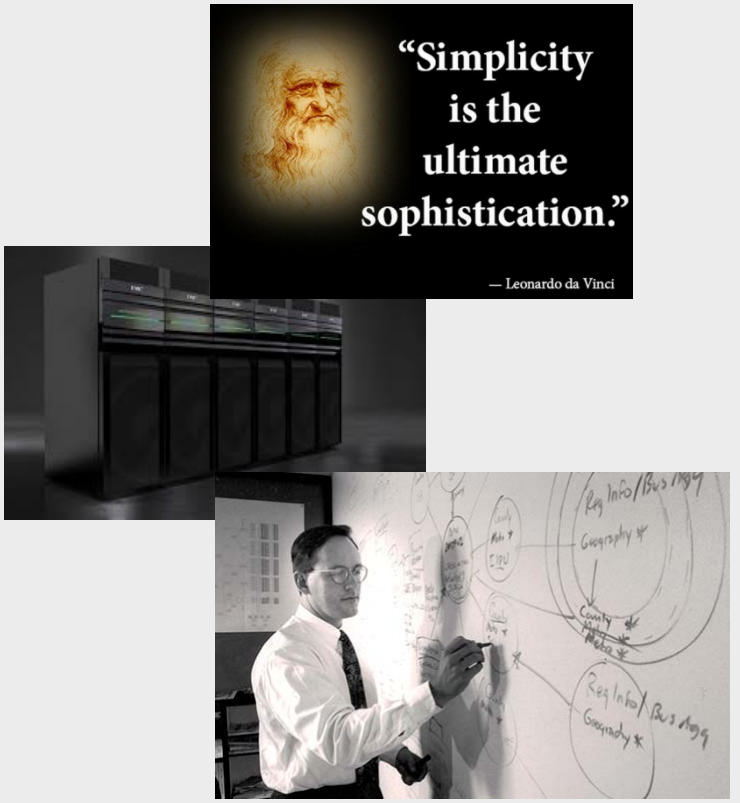
Centraal Bureau voor de Statistiek

Statistics Netherlands and Big Data
Why a Big Data approach?

- Shorter time to publication
- Respond to current events
- Higher reliability
- More detail
- More efficient processes

Considerations:

- Infrastructure
- Competences
- Culture

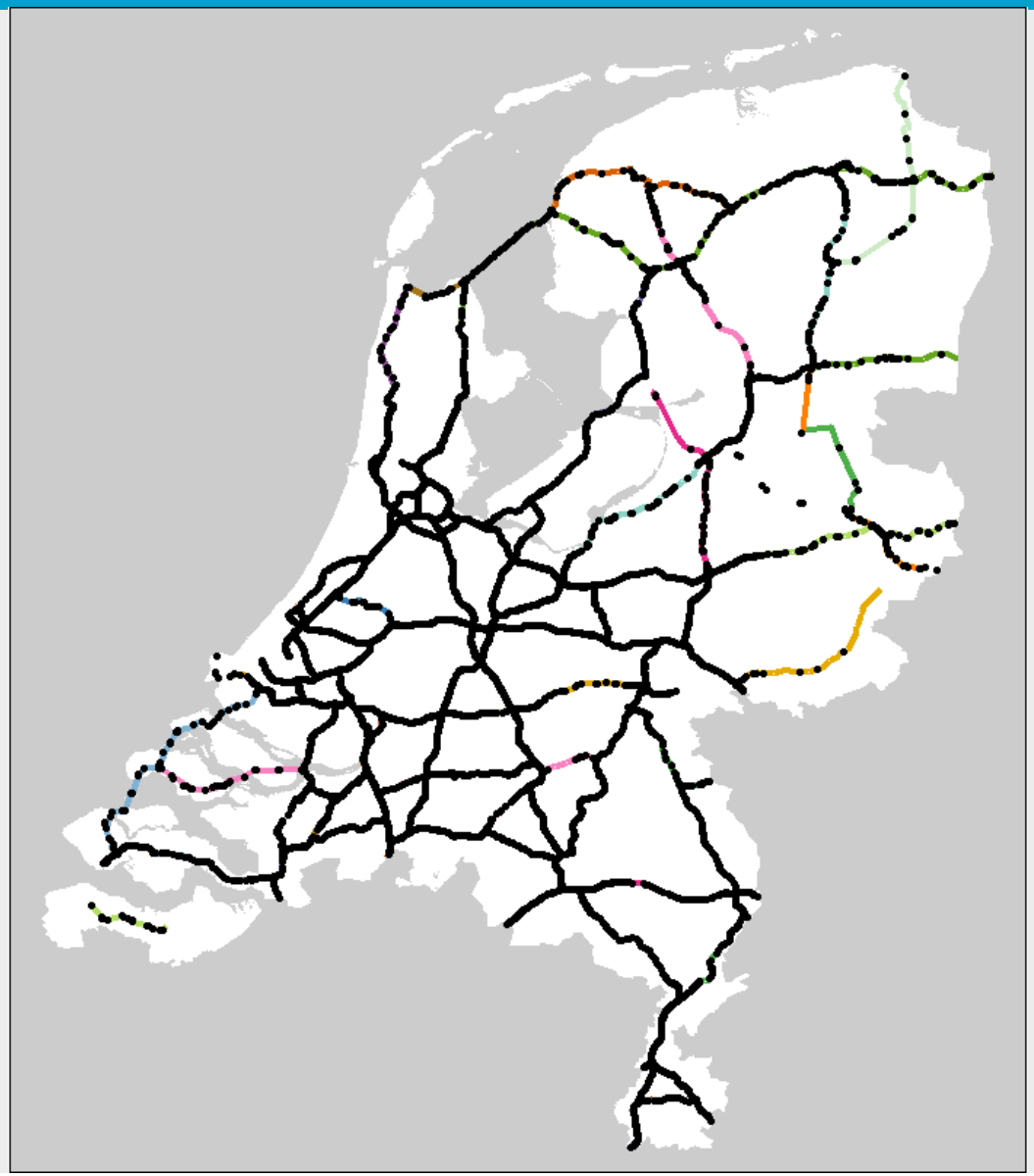


Road sensors

Road sensor data

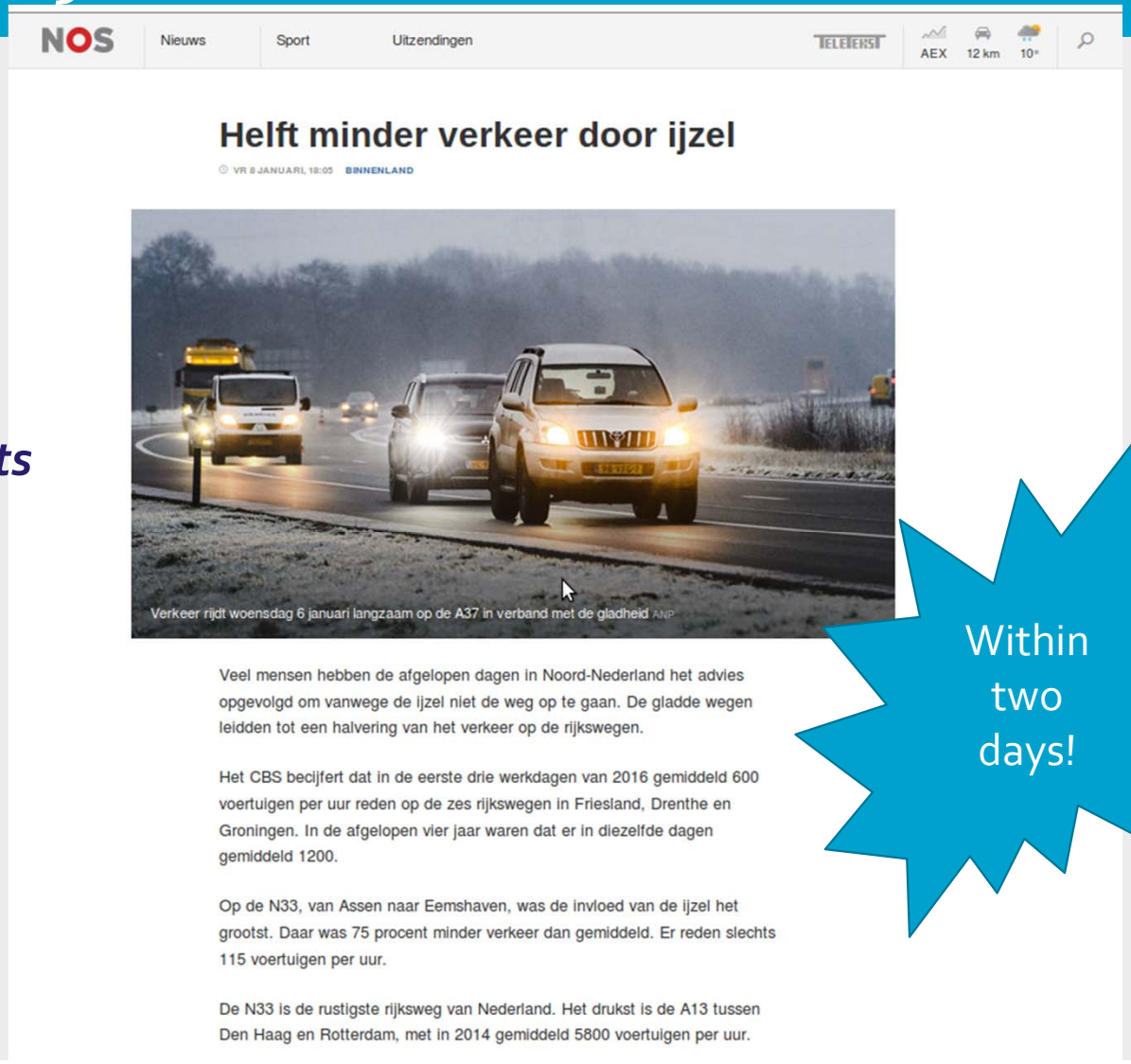
- Passing vehicle counts for each minute (24/7) at about 60.000 sensors in the Netherlands
- Types of sensors:
  - Induction loop
  - Camera
  - Bluetooth
- Length categories (e.g. small, medium, long vehicles)
- Large volume: approx. 230 mln records/day

Dutch highways with road sensors



Data journalism and (almost) real time statistics

Respond to *current events*



NOS Nieuws Sport Uitzendingen TELETEKST AEX 12 km 10°

Helft minder verkeer door ijzel

VR 6 JANUARI, 18:05 BINNENLAND

Verkeer rijdt woensdag 6 januari langzaam op de A37 in verband met de gladheid ANP

Veel mensen hebben de afgelopen dagen in Noord-Nederland het advies opgevolgd om vanwege de ijzel niet de weg op te gaan. De gladde wegen leidden tot een halvering van het verkeer op de rijkswegen.

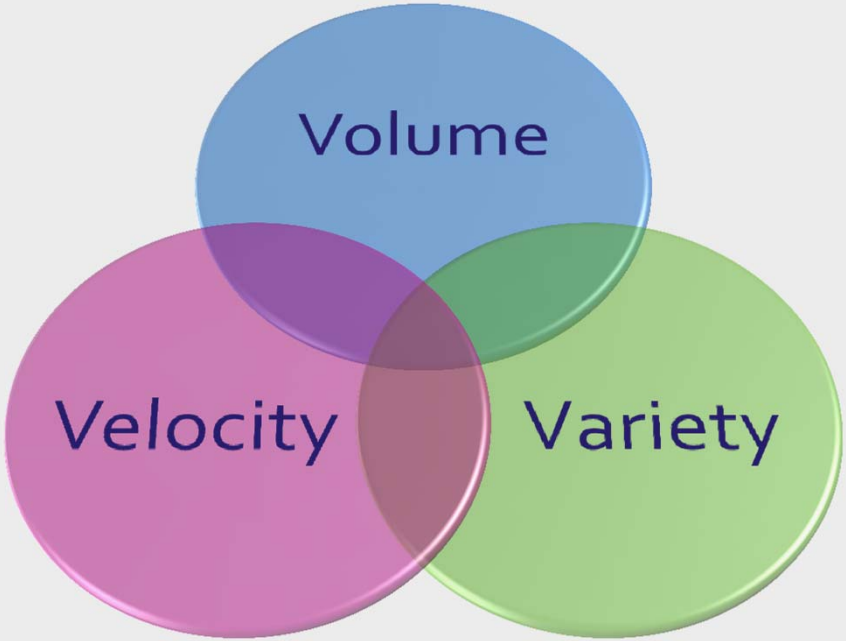
Het CBS becijfert dat in de eerste drie werkdagen van 2016 gemiddeld 600 voertuigen per uur reden op de zes rijkswegen in Friesland, Drenthe en Groningen. In de afgelopen vier jaar waren dat er in diezelfde dagen gemiddeld 1200.

Op de N33, van Assen naar Eemshaven, was de invloed van de ijzel het grootst. Daar was 75 procent minder verkeer dan gemiddeld. Er reden slechts 115 voertuigen per uur.

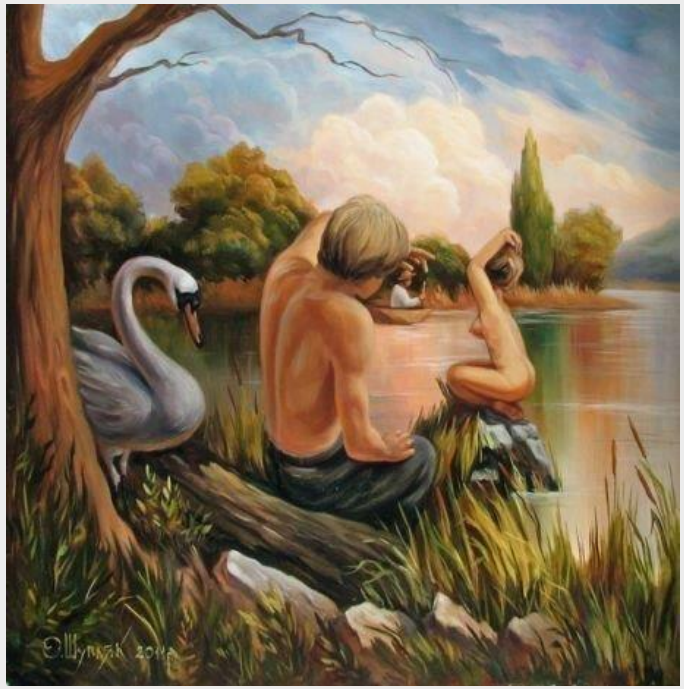
De N33 is de rustigste rijksweg van Nederland. Het drukst is de A13 tussen Den Haag en Rotterdam, met in 2014 gemiddeld 5800 voertuigen per uur.

Within two days!

What is Big Data?



The signal and the Data

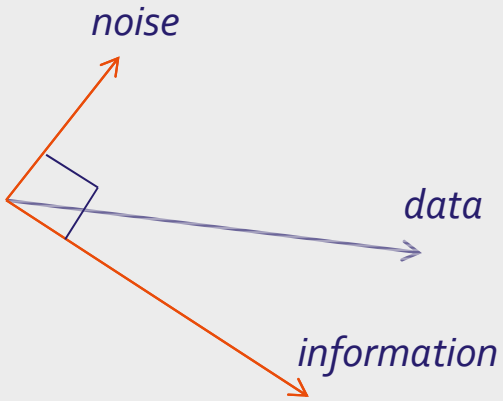


Oleg Shuplyak

The Signal and the Data

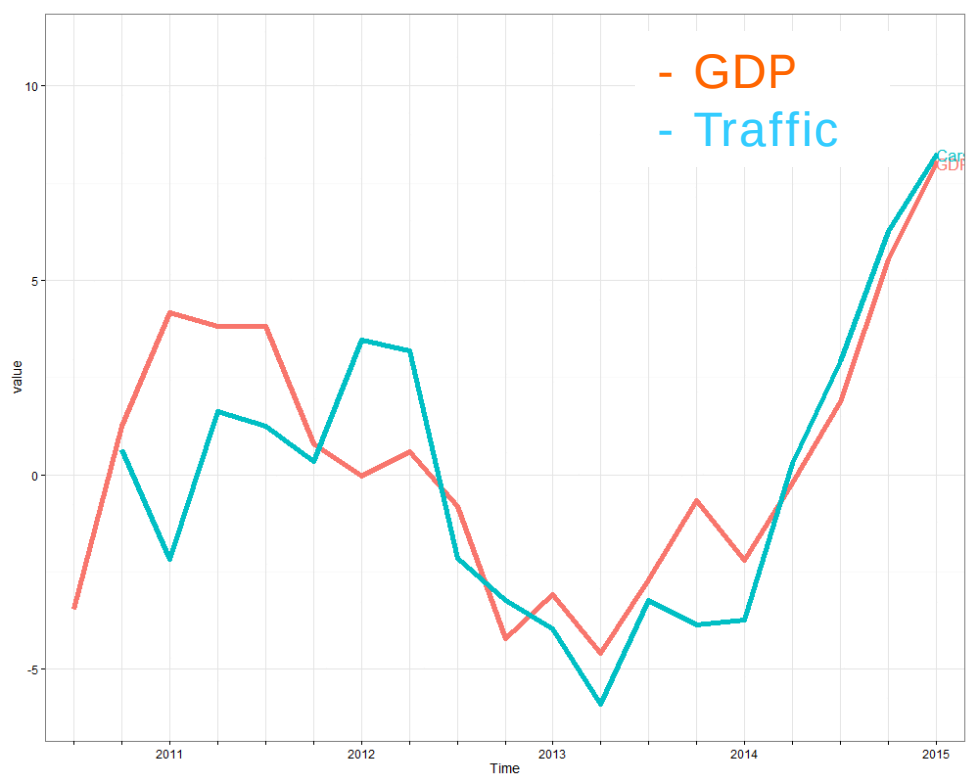
Tuning the data

data = information + noise



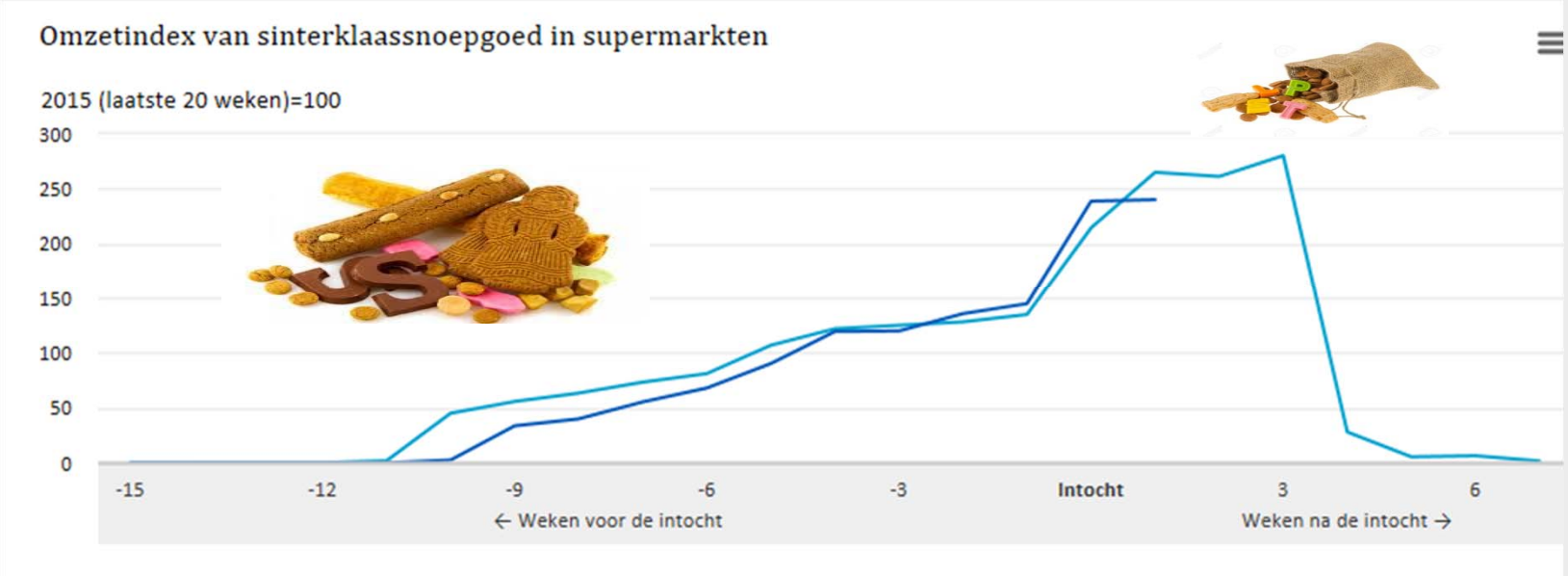
Noise is that part of the data that is not relevant!

GDP and Road Traffic Data

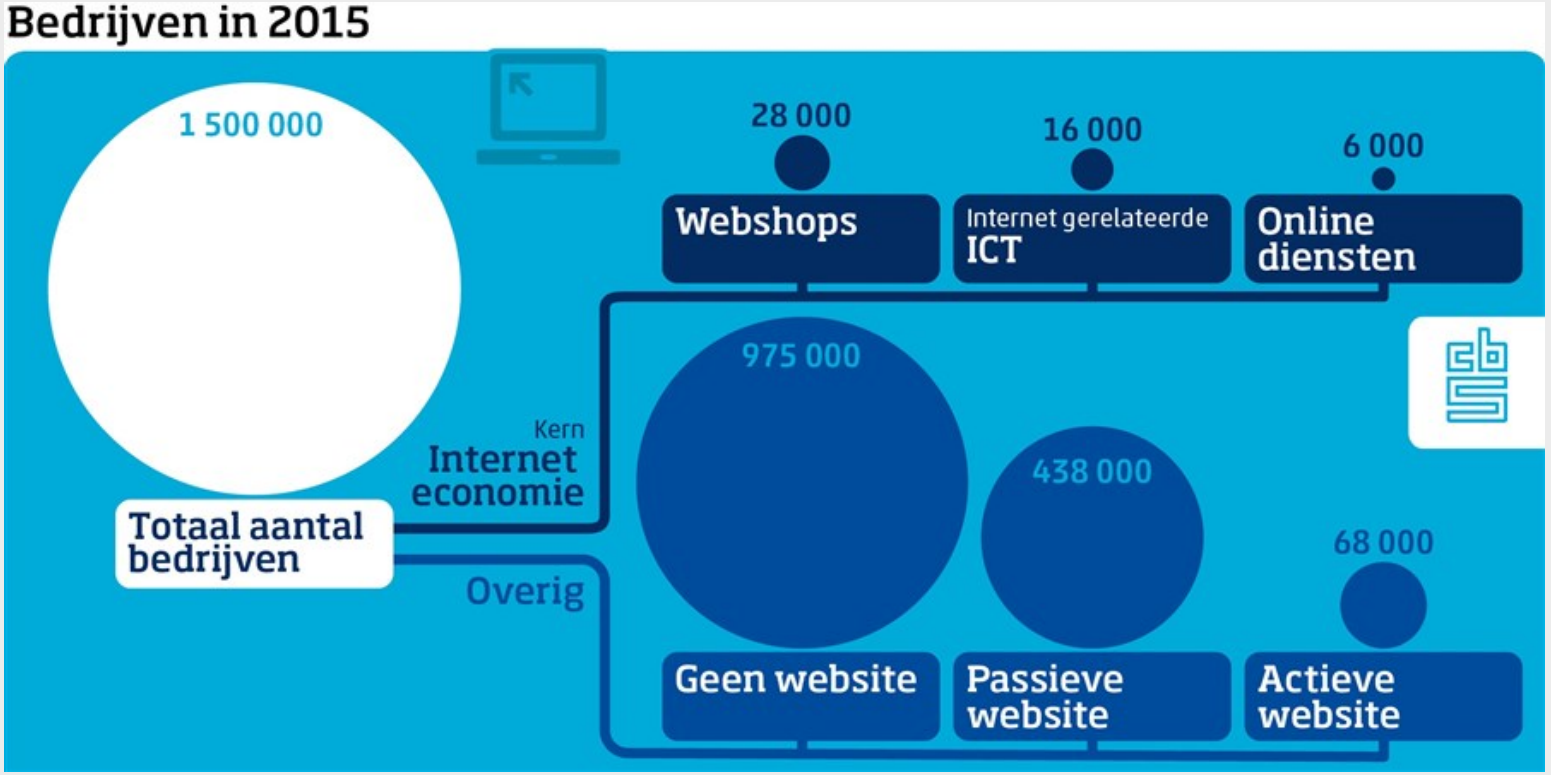


Scanner data from supermarkets

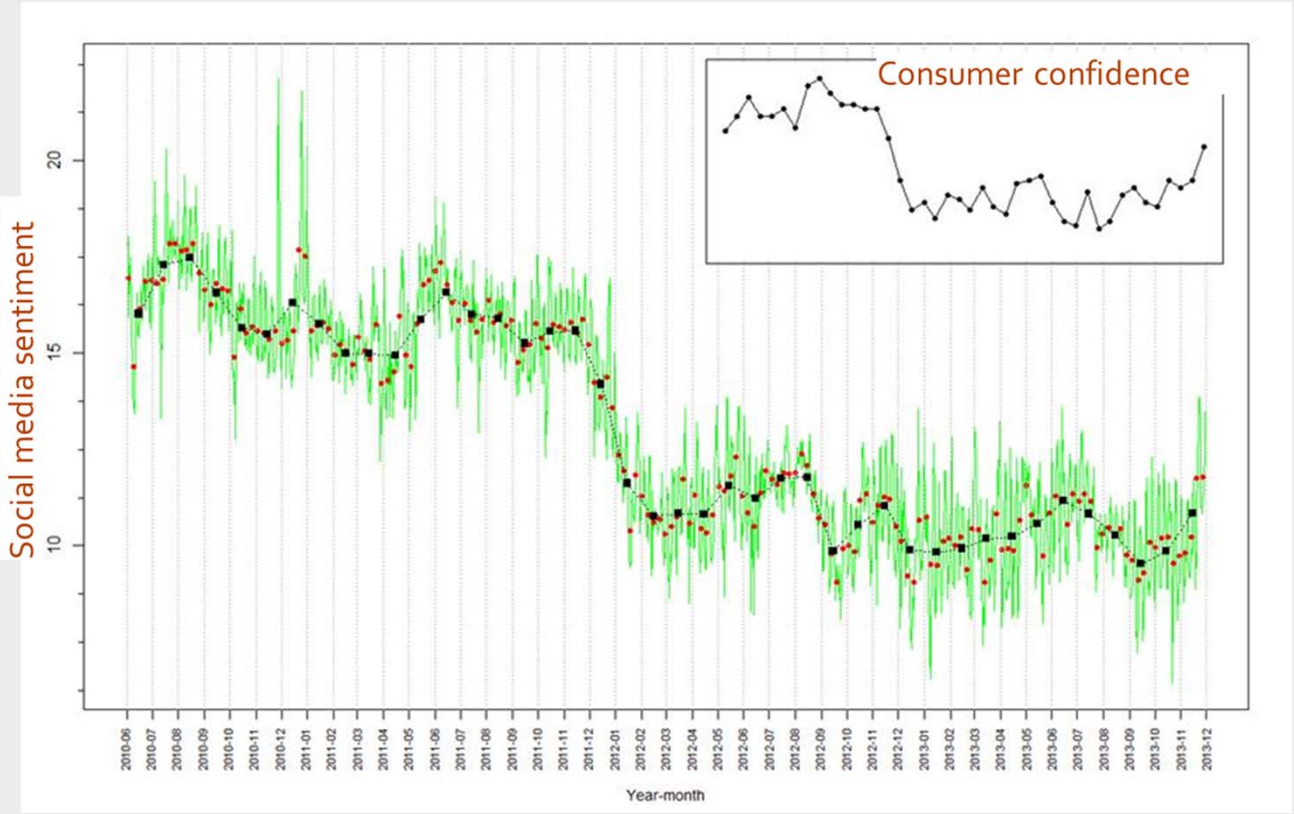
Scanner Data: the gingerbread index



Web scraped data



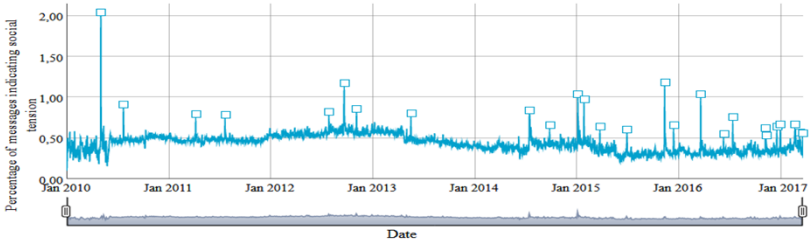
Sentiment on Social Media



Sentiment based social tension indicator

Social tension indicator based on social media

How can social mec
Statistics Netherlan
indicator. The indic
measurements of p
was conducted to c
The indicator identi

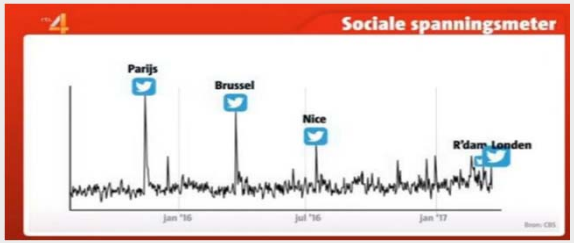


High degree of social tension after Dam Screamer incident and terrorist attacks

This visualisation shows peaks in the social tension indicator. A larger number of messages were posted on or just after the days on which incidents took place which created feelings of unsafety and unrest. The big spike in 2010, for example, is related to the disruption of national Remembrance Day commemorations by the 'Dam Screamer' at Amsterdam Dam Square on 4 May. People's responses to terrorist attacks are also reflected in the social tension indicator: the terrorist attacks in Paris (13 November 2015) and Brussels (22 March 2016) caused a peak in tensions in the Netherlands. In addition, the MH17 disaster (17 July 2014) resulted in strong feelings of unsafety and unrest. Other types of events such as the election of Donald Trump in the United States on 9 November 2016 caused social tensions as well.

Social tensions slightly down in recent years

According to the indicator, overall social tension increased slightly between 2010 and 2013, followed by a slight decrease; this trend can be compared with the results from the annual Safety Monitor, for which a representative group



https://www.cbs.nl/en-gb/our-services/innovation/project/social-tension-indicator-based-on-social-media

COOSTO

Aerial photos and solar panels

Conclusion

- Big data is normally tuned to the problems we need to answer
- Different big data sources used by CBS:
  - Road sensor data
  - Social media data
  - Scanner data
  - Web scraped data
  - Aerial photo‘s
  - AIS
- Sometimes we reach the news with our big data projects