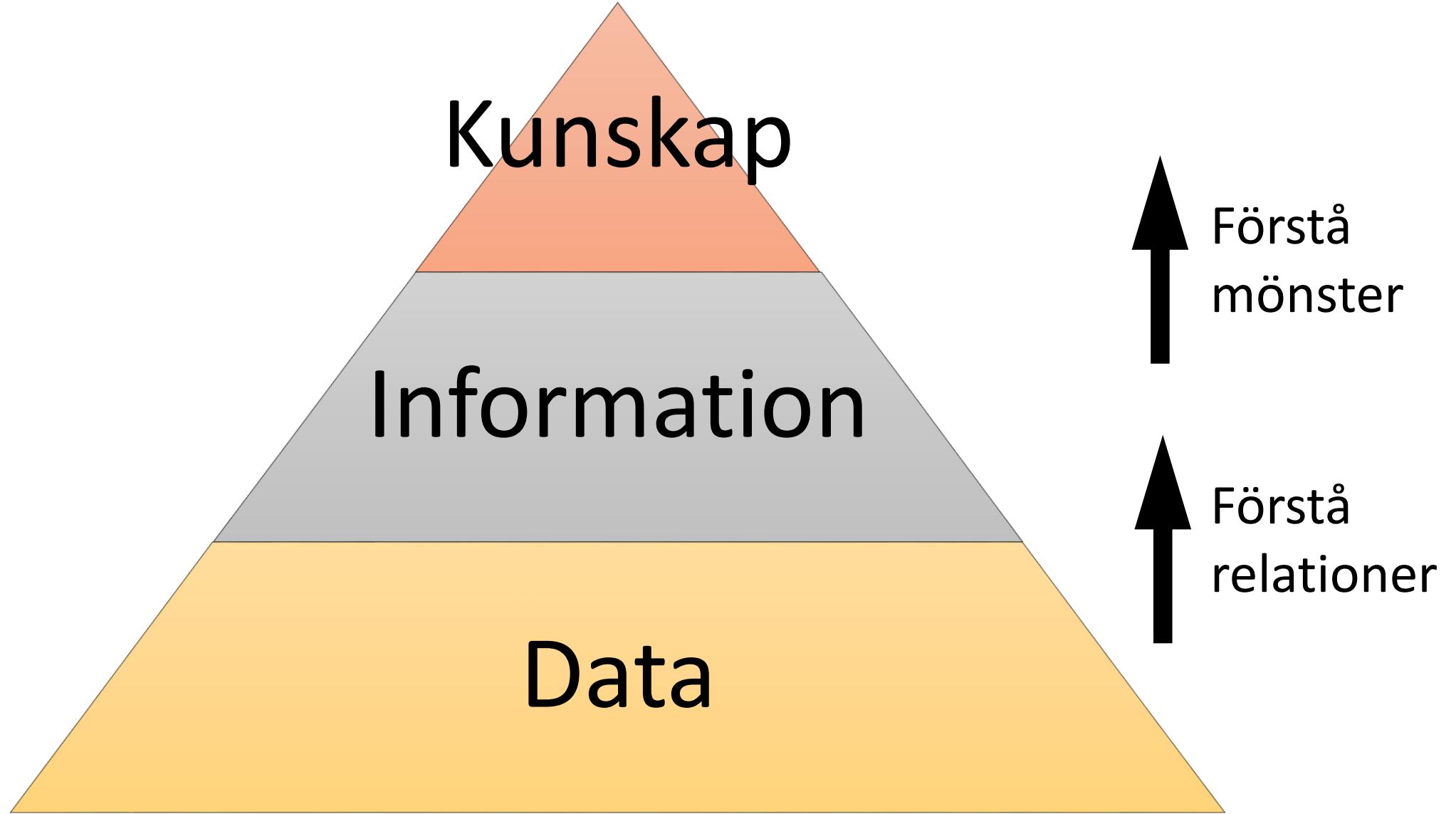


# Data, Information, Kunskap

Peter Krantz, Avdelningen för digitalisering, SKL



Se t.ex. DIKW [https://en.wikipedia.org/wiki/DIKW\\_Pyramid](https://en.wikipedia.org/wiki/DIKW_Pyramid)

Vi börjar i rymden



NEMCO





Kostnad: ca 2.5 mdr kr

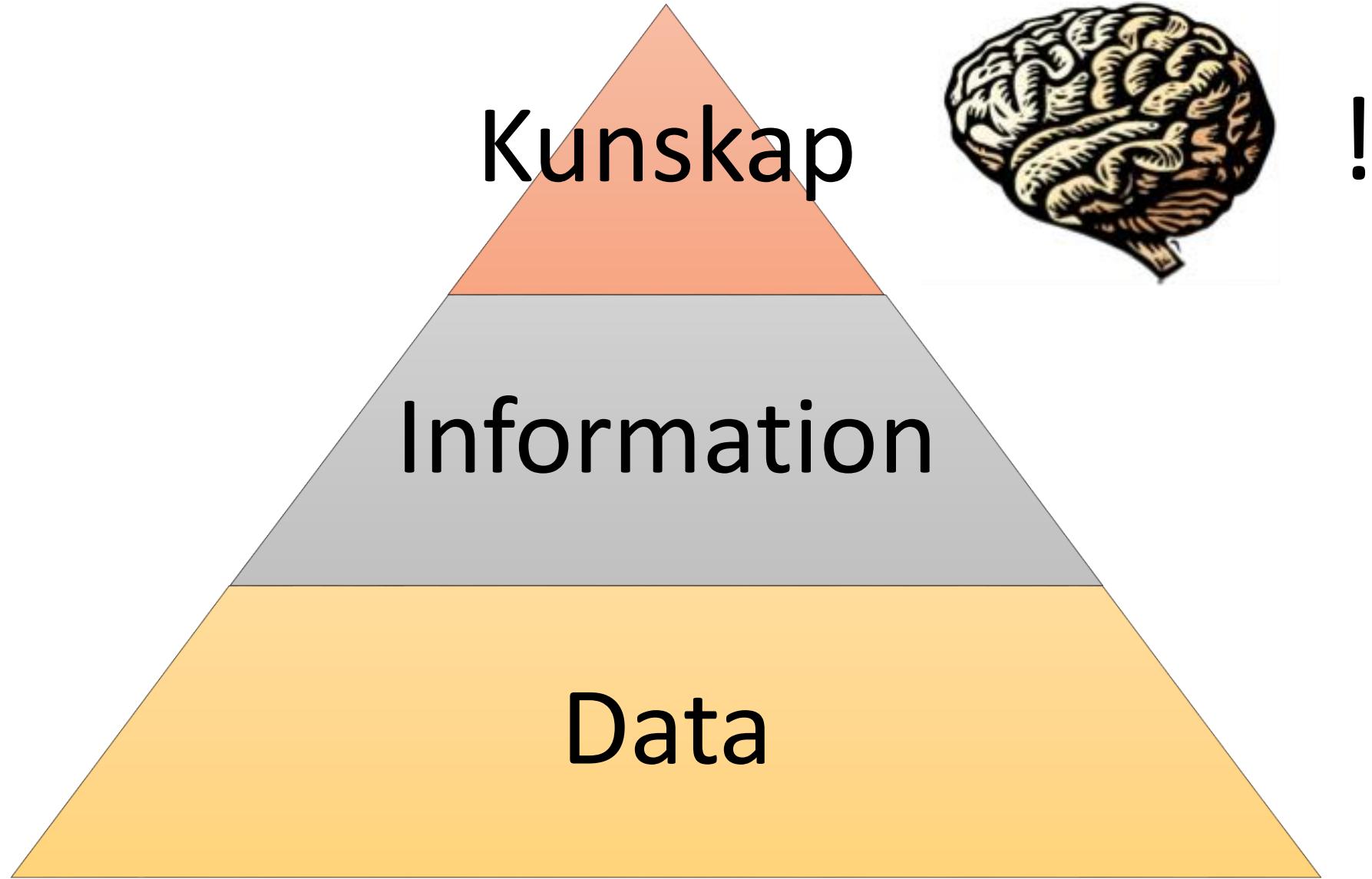


Varför jobbar vi med information och  
visualisering?

# Data som förutsättning för kunskap

Med fri tillgång till data maximeras möjligheterna för andra att skapa information som underlag för kunskap

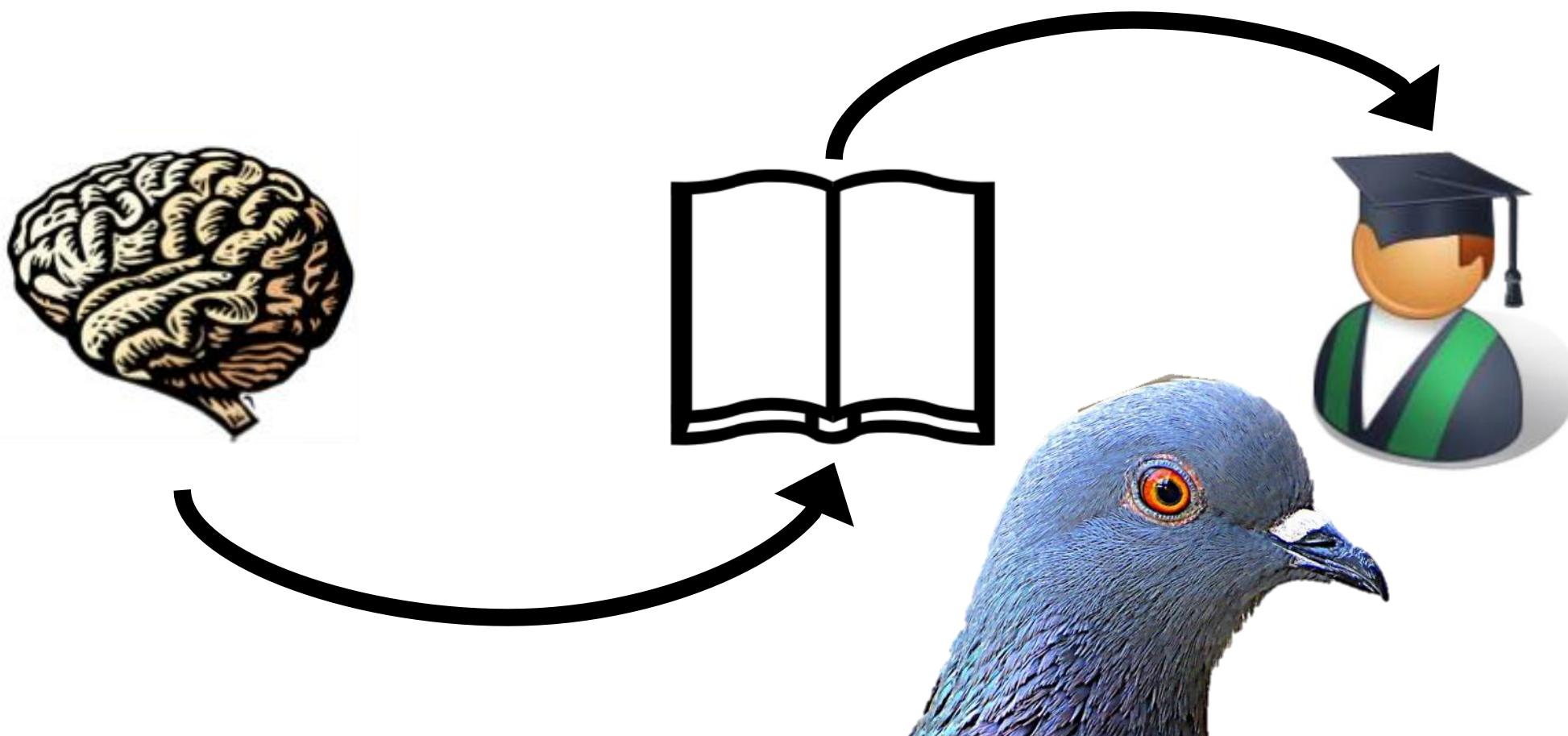
...och med kunskap kan vi förändra samhället på ett faktabaserat sätt.

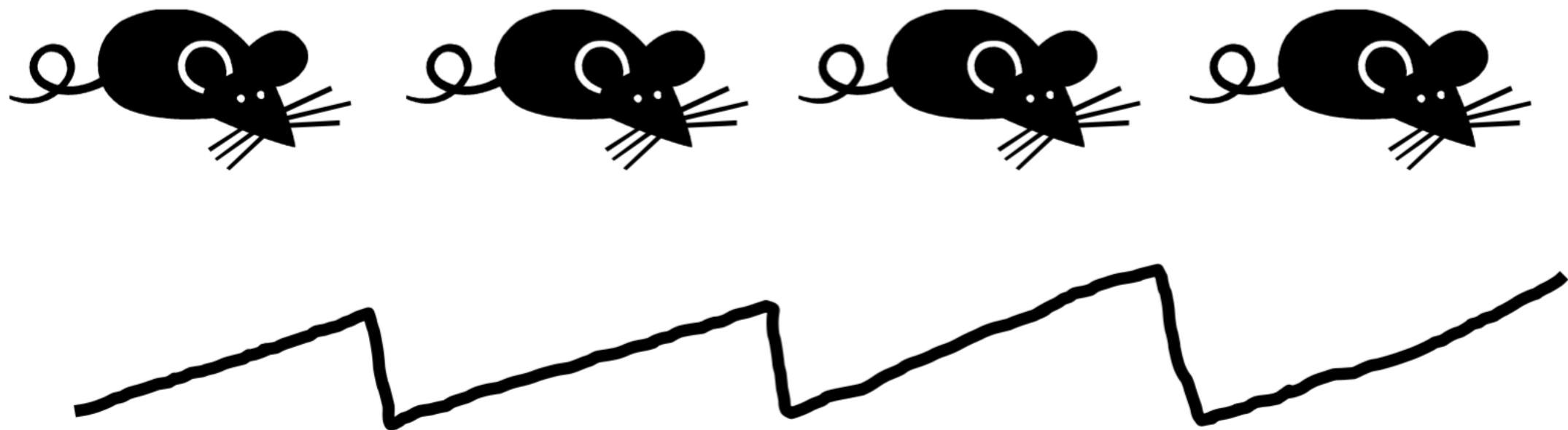


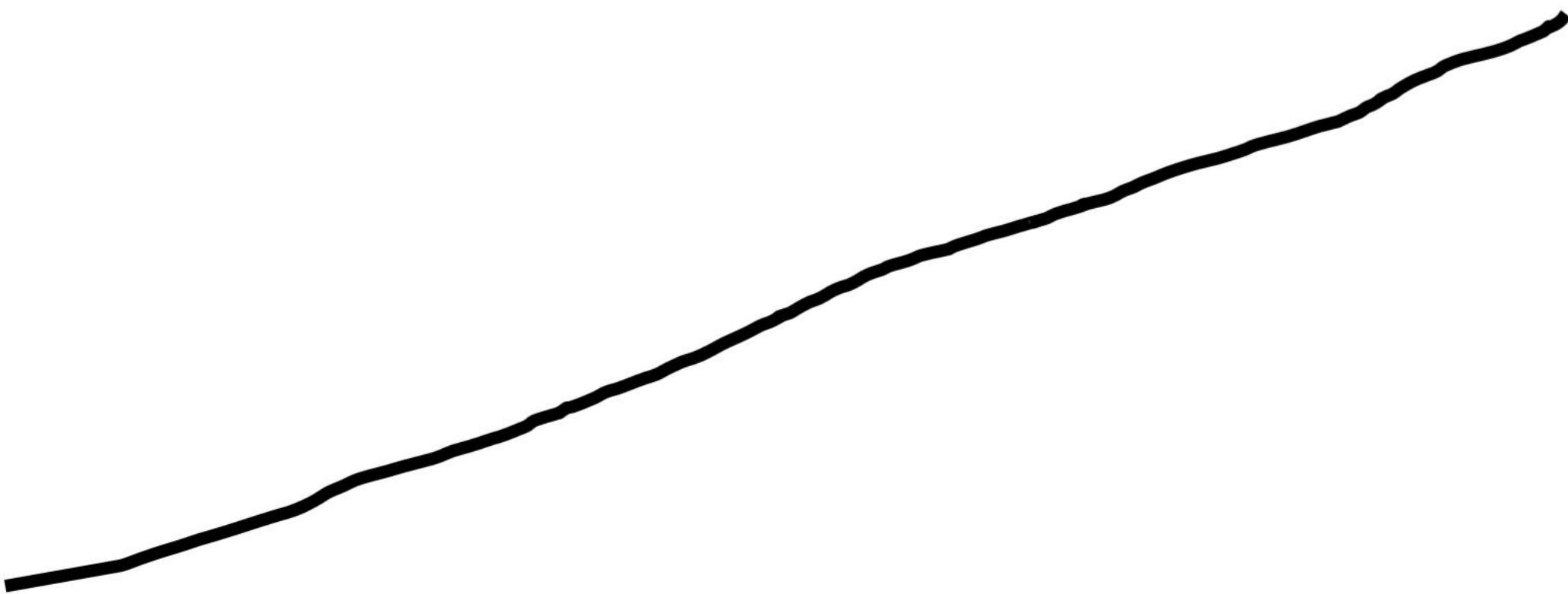
Se t.ex. DIKW [https://en.wikipedia.org/wiki/DIKW\\_Pyramid](https://en.wikipedia.org/wiki/DIKW_Pyramid)

# Unikt för människan?

Kunskapsuttryck som överlever generationer och avstånd





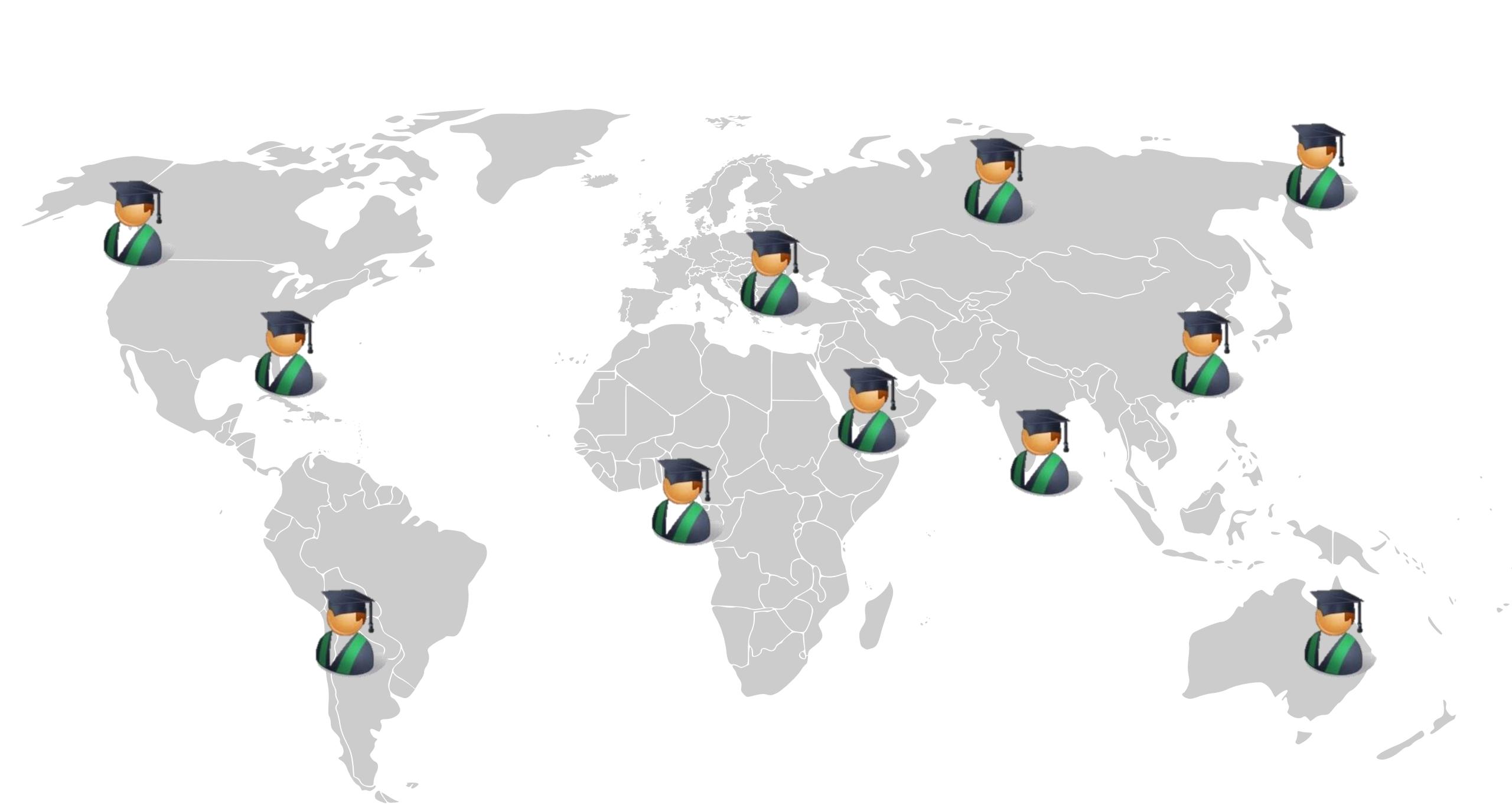


# Internet

Unik egenskap med digital information:

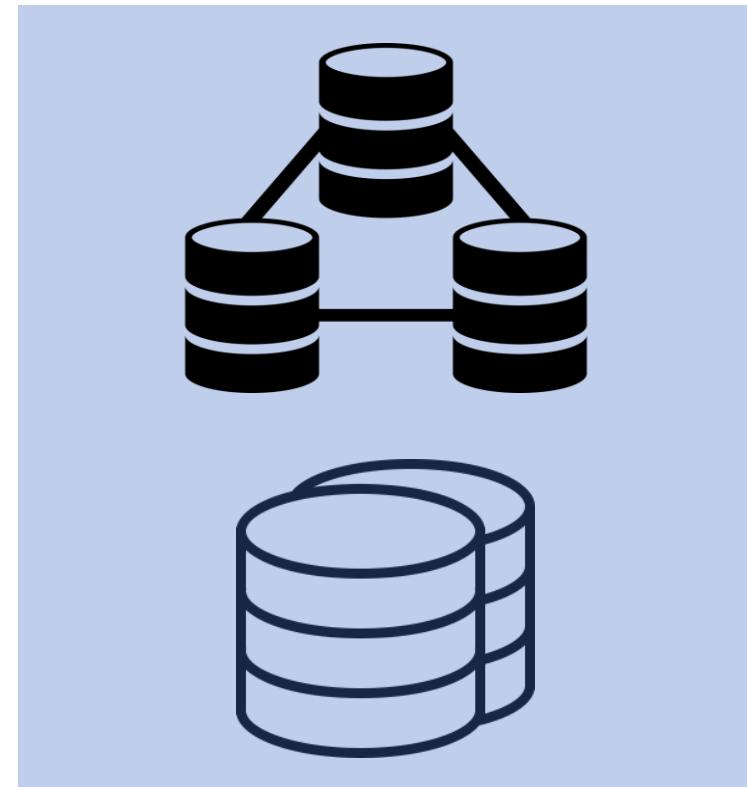
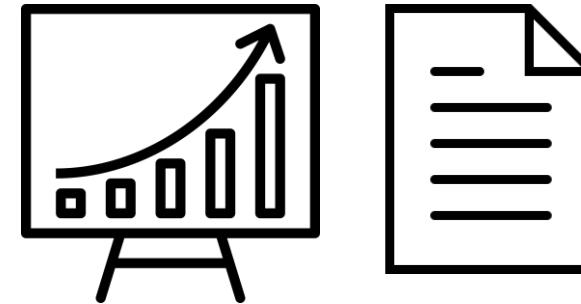
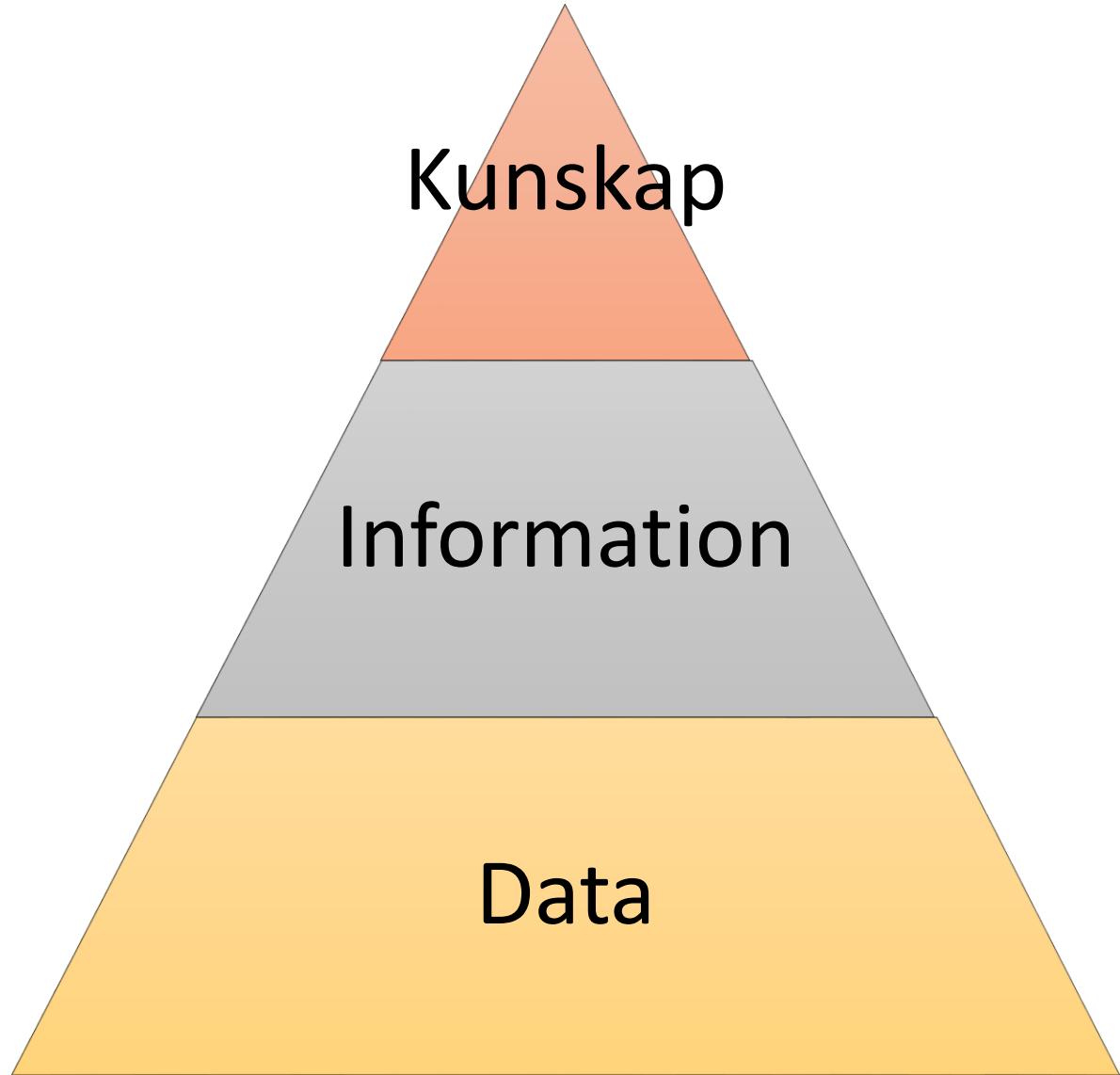
Vi kan dela med oss av den,

...och fortfarande ha den kvar!



Kan vi bli smartare med globalt utbyte av  
information?

Hur maximerar vi nyttan av information?



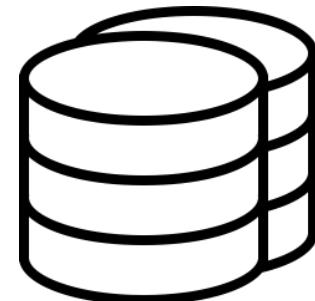
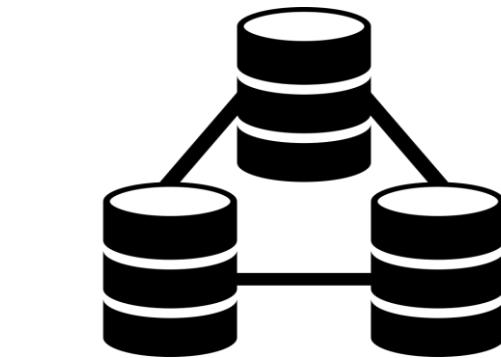
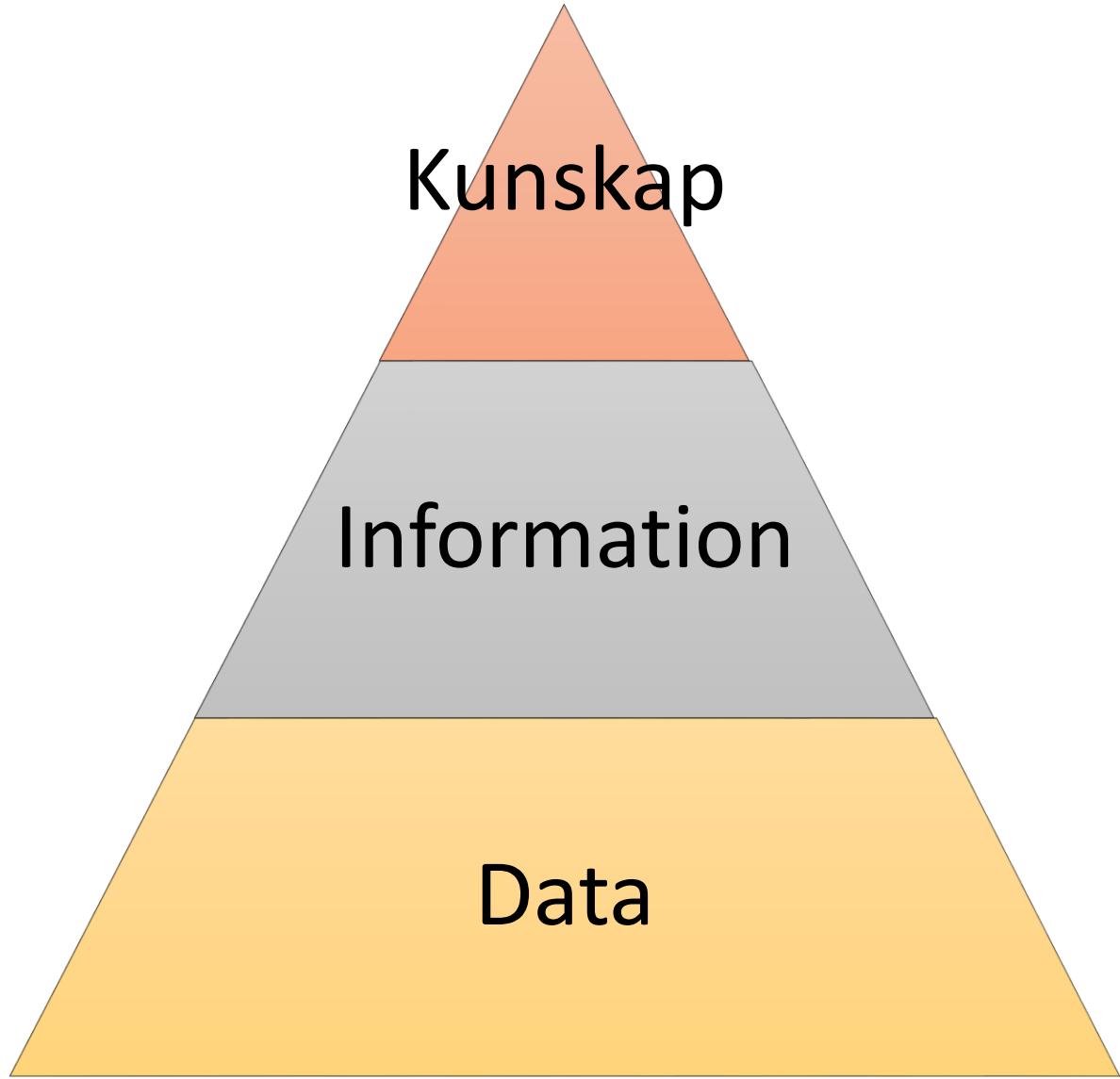
# Några kvalitetsaspekter

1. Öppenhet
2. Hög upplösning
3. Återanvändbarhet (standardisering och interoperabilitet)
4. Begriplighet

# 1. Öppenhet: Öppna data

“Data kan kallas öppna data om vem som helst fritt får använda, återanvända och distribuera dessa med som största motprestation att ange källa eller krav på att dela data på samma sätt.”

[opendefinition.org](https://opendefinition.org)

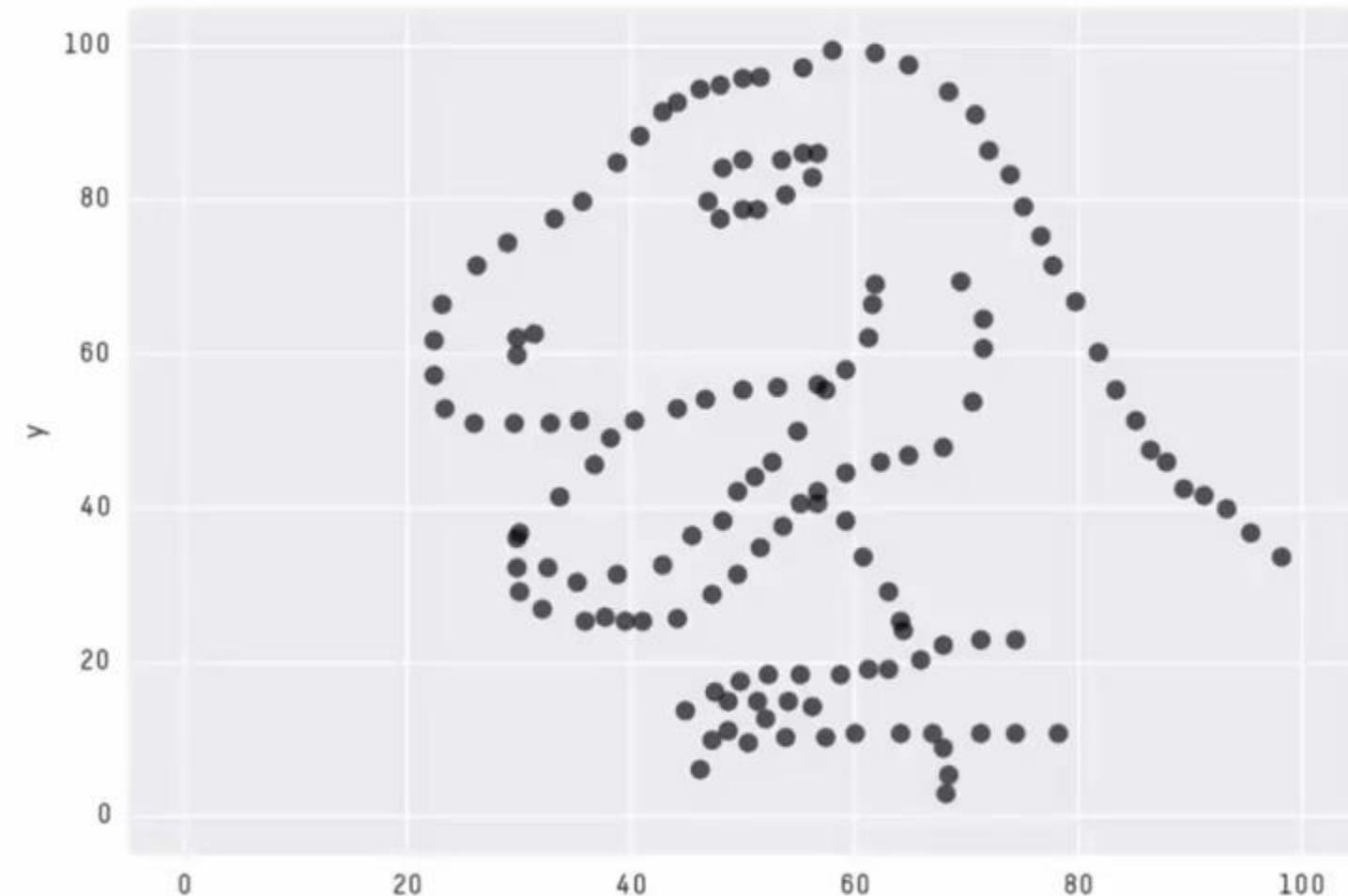


Öppet licensierad standardiserad information

## 2. Upplösning

"Medelvärdet för X var 54.26 förra året"

Kan man förstå mer om informationen har högre upplösning?



X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526

### 3. Återanvändbarhet (standardisering)

Om informationen är standardiserad och länkad med annan information?

”Vilka professionella fotbollsspelare, exakt 170cm långa, har spelat i ett lag som har ett kontor/säte i kommuner med minst 2000 anställda?”

Wikidata Query

Exempel Hjälp Verktyg svenska

### Frågehjälpare

|                |                               |                            |  |  |
|----------------|-------------------------------|----------------------------|--|--|
| item           | instans av                    | människa                   |  |  |
| item           | sysselsättning                | fotbollsspelare            |  |  |
| mp             | instans av<br>undergrupp till | någon<br>Sveriges kommuner |  |  |
| item           | medlem i idrottslag           | team                       |  |  |
| item           | höjd                          | height                     |  |  |
| team           | säte                          | headquarterloc             |  |  |
| headquarterloc | inom administrativt område    | mp                         |  |  |
| mp             | folkmängd                     | population                 |  |  |
| mp             | antal anställda               | employeecount              |  |  |

+ Filtrera

+ Visa

Gräns 50

```

1 PREFIX wd: <http://www.wikidata.org/entity/>
2 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
3 PREFIX wikibase: <http://wikiba.se/ontology#>
4 PREFIX schema: <http://schema.org/>
5 PREFIX bd: <http://www.bigdata.com/rdf#>
6 SELECT DISTINCT ?item ?itemLabel ?mpLabel ?teamLabel ?height ?population ?employeecount ?headquarterlocLabel
7 WHERE {
8   ?item wdt:P31 wd:Q5 .
9   ?item wdt:P106 wd:Q937857 .
10  ?item wdt:P54 ?team .
11  ?item wdt:P2048 ?height .
12  ?team wdt:P159 ?headquarterloc .
13  ?headquarterloc wdt:P131 ?mp .
14  ?mp wdt:P31/wdt:P279* wd:Q127448 .
15  ?mp wdt:P1082 ?population .
16  ?mp wdt:P1128 ?employeecount .
17  FILTER(?employeecount > 2000 && ?height = 170).
18  #FILTER(?height >= 190).
19  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,fr" }
20 }
21 ORDER BY DESC(?height)
22 LIMIT 50

```

| item                           | itemLabel             | mpLabel          | teamLabel               | height | population | employeecount |
|--------------------------------|-----------------------|------------------|-------------------------|--------|------------|---------------|
| <a href="#">Q wd:Q22282555</a> | Lumala Abdu           | Kalmar kommun    | Kalmar FF               | 170    | 67528      | 6775          |
| <a href="#">Q wd:Q491163</a>   | Anders Linderoth      | Växjö kommun     | Östers IF               | 170    | 91251      | 7825          |
| <a href="#">Q wd:Q658279</a>   | Paulo José Figueiredo | Växjö kommun     | Östers IF               | 170    | 91251      | 7825          |
| <a href="#">Q wd:Q3733085</a>  | Espen Minde           | Bodens kommun    | Bodens BK               | 170    | 28184      | 3075          |
| <a href="#">Q wd:Q3423430</a>  | Reggie Lambe          | Nyköpings kommun | Nyköpings BIS           | 170    | 55552      | 5075          |
| <a href="#">Q wd:Q349524</a>   | Rune Börjesson        | Göteborgs kommun | Örgryte IS              | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q948874</a>   | Lieke Martens         | Göteborgs kommun | Kopparbergs/Göteborg FC | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q592622</a>   | Ghandi Kassenu        | Göteborgs kommun | BK Häcken               | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q444055</a>   | Johanna Almgren       | Göteborgs kommun | Kopparbergs/Göteborg FC | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q2678783</a>  | John Owoeri           | Göteborgs kommun | BK Häcken               | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q2861144</a>  | Ari Freyr Skúlason    | Göteborgs kommun | BK Häcken               | 170    | 565496     | 34725         |
| <a href="#">Q wd:Q349524</a>   | Rune Börjesson        | Mölndals kommun  | Örgryte IS              | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q444055</a>   | Johanna Almgren       | Mölndals kommun  | Kopparbergs/Göteborg FC | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q2678783</a>  | John Owoeri           | Mölndals kommun  | BK Häcken               | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q2861144</a>  | Ari Freyr Skúlason    | Mölndals kommun  | BK Häcken               | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q592622</a>   | Ghandi Kassenu        | Mölndals kommun  | BK Häcken               | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q948874</a>   | Lieke Martens         | Mölndals kommun  | Kopparbergs/Göteborg FC | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q512906</a>   | Stéphanie Öhrström    | Mölndals kommun  | Jitex BK                | 170    | 66733      | 4825          |
| <a href="#">Q wd:Q349524</a>   | Rune Börjesson        | Partille kommun  | Örgryte IS              | 170    | 37931      | 2875          |

# Standardisering av informasjon



Hur ser det ut i it-området då?



*"The output from the SM\_FORCES application code as required by a MSOP Project Software Interface Specification (SIS) was to be in **metric units of Newton-seconds** (N-s).*

*Instead, the data was reported in **English units of pound-seconds** (lbf-s)."*

[NASA report on MCO vehicle loss](#)



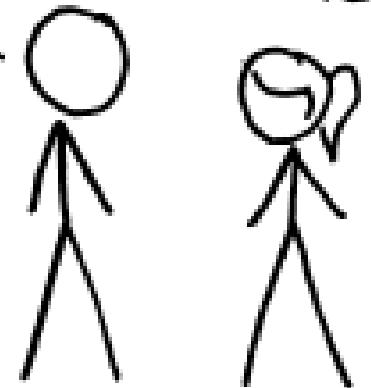
# Standardising

HOW STANDARDS PROLIFERATE:  
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:  
THERE ARE  
14 COMPETING  
STANDARDS.

14?! RIDICULOUS!  
WE NEED TO DEVELOP  
ONE UNIVERSAL STANDARD  
THAT COVERS EVERYONE'S  
USE CASES.

YEAH!



Soon:

SITUATION:  
THERE ARE  
15 COMPETING  
STANDARDS.

# Interoperabilitet

”förmågan hos olika system, ofta i datorsammanhang, att fungera tillsammans och kunna kommunicera med varandra”

# Interoperabilitet

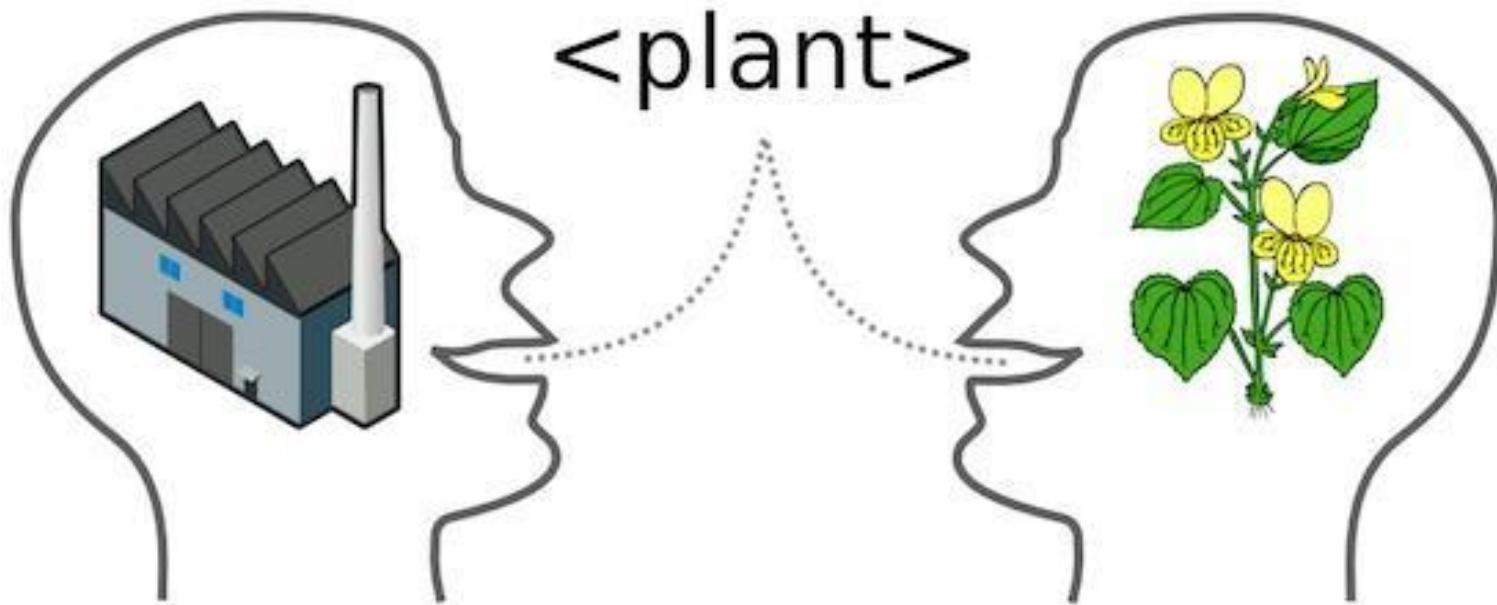
Rättslig

Organisatorisk

Semantisk

Teknisk

# Semantisk interoperabilitet



Men hur blir det kunskap?

## 4. Begriplighet

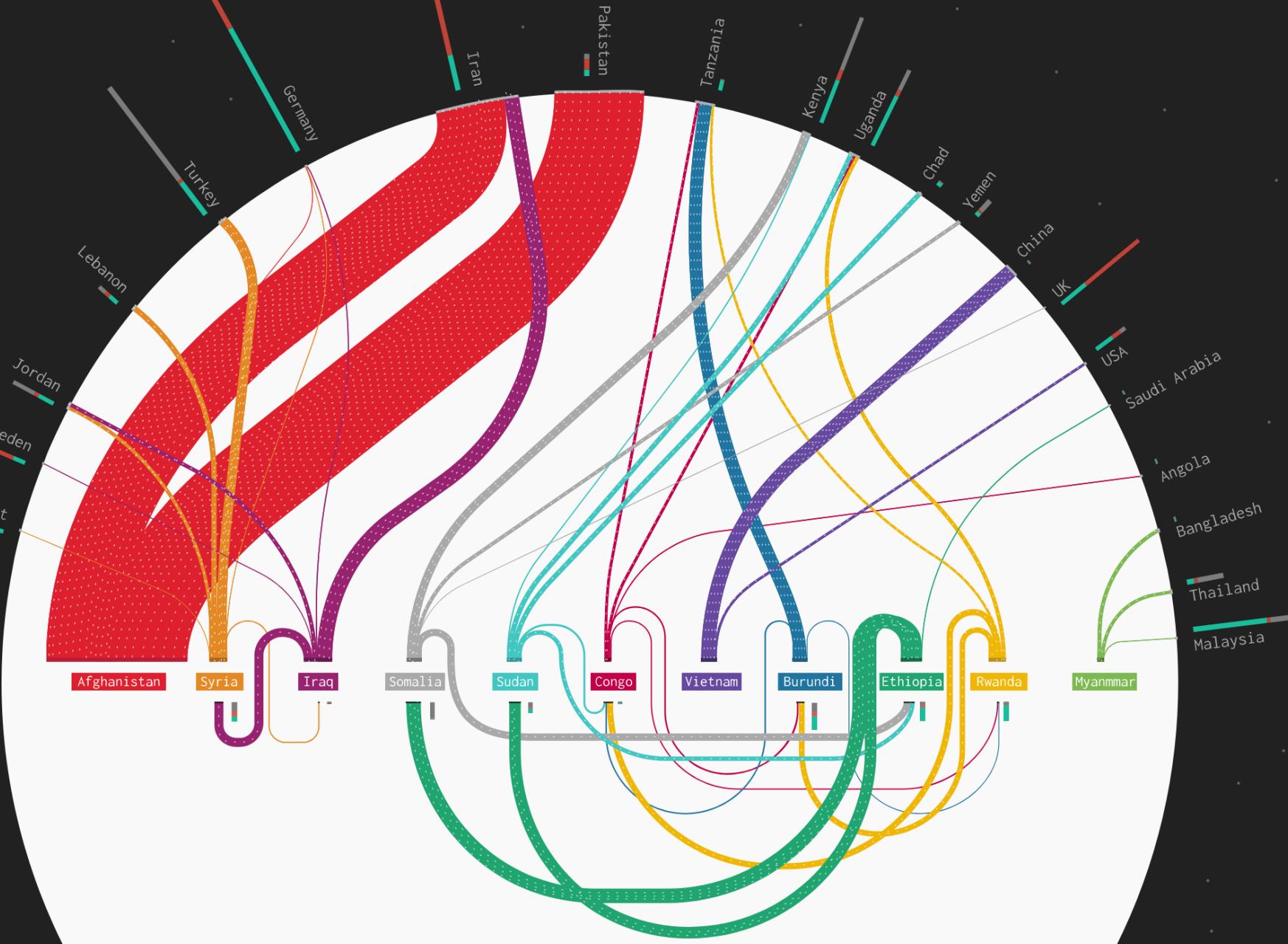
För att förstå information behöver vi uppfatta den med våra sinnen

t.ex. genom att visualisera den...

# Exodus

Uncover the causalities behind countries with the largest refugee-movements.

This dataviz shows the total number of refugees that have migrated from their homes to new lands, from the countries with a major refugee-crisis, from 1960s.



Every country has a story that lead to these current circumstances.

explore more

# Titanic Survivors

**Survived**

Survived

Perished

**Sex**

Female

Male

**Age**

Child

Adult

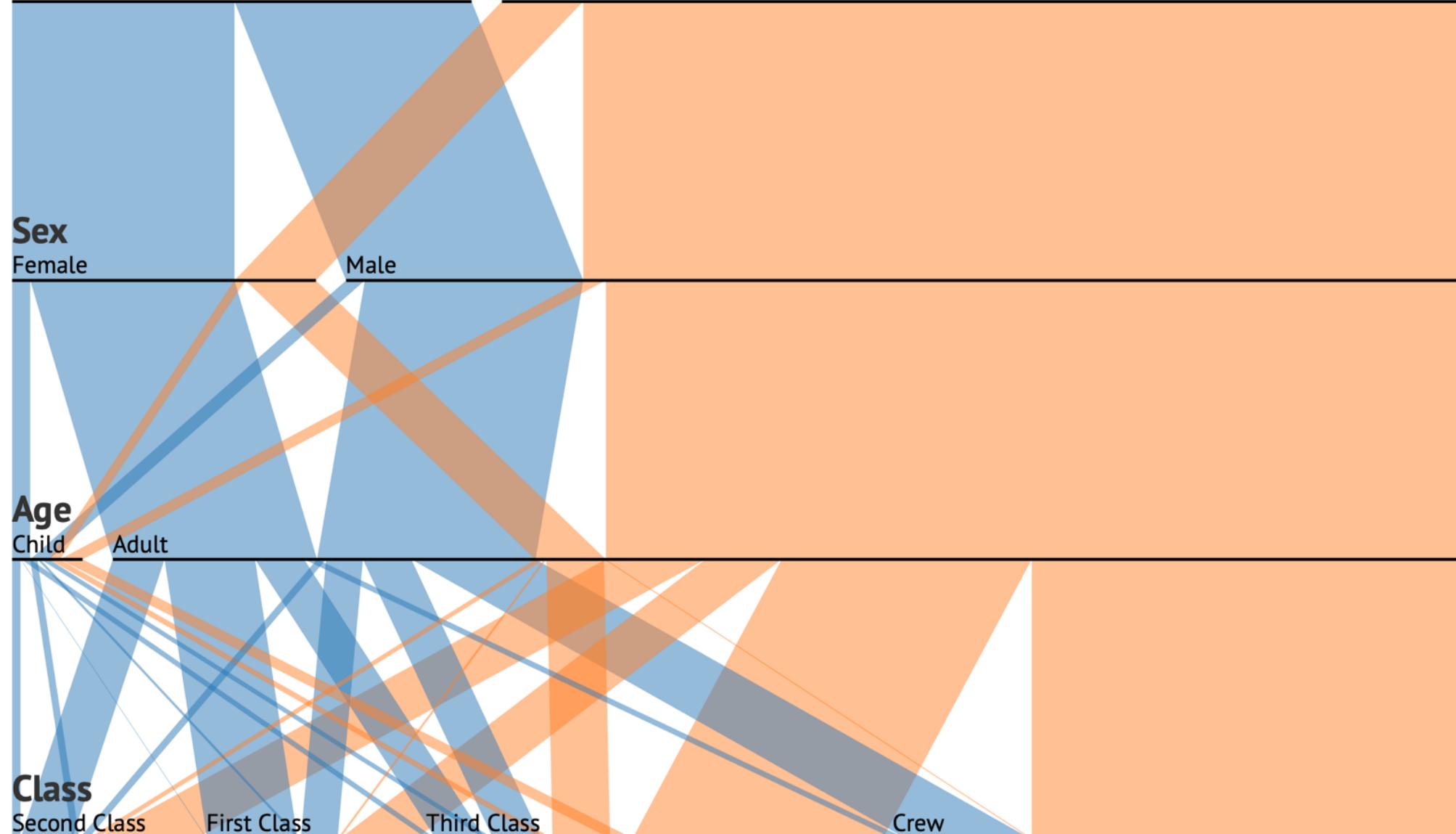
**Class**

Second Class

First Class

Third Class

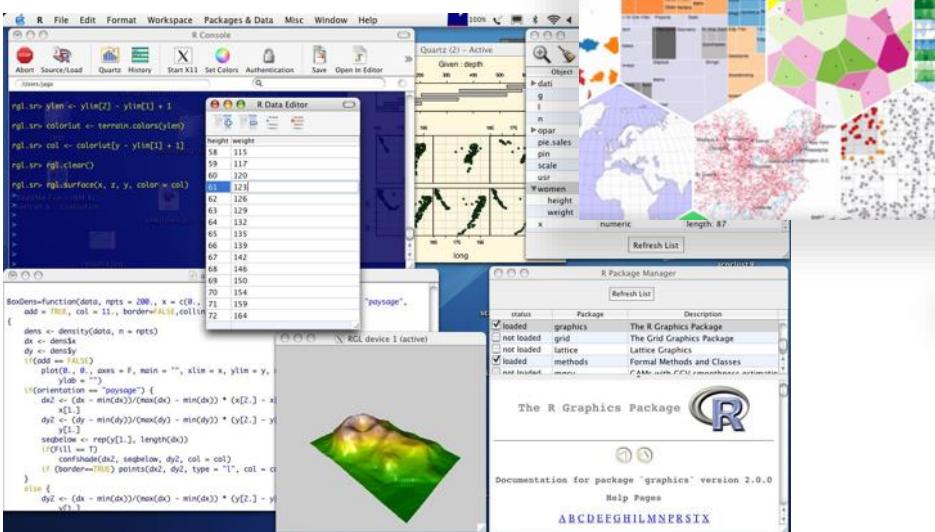
Crew



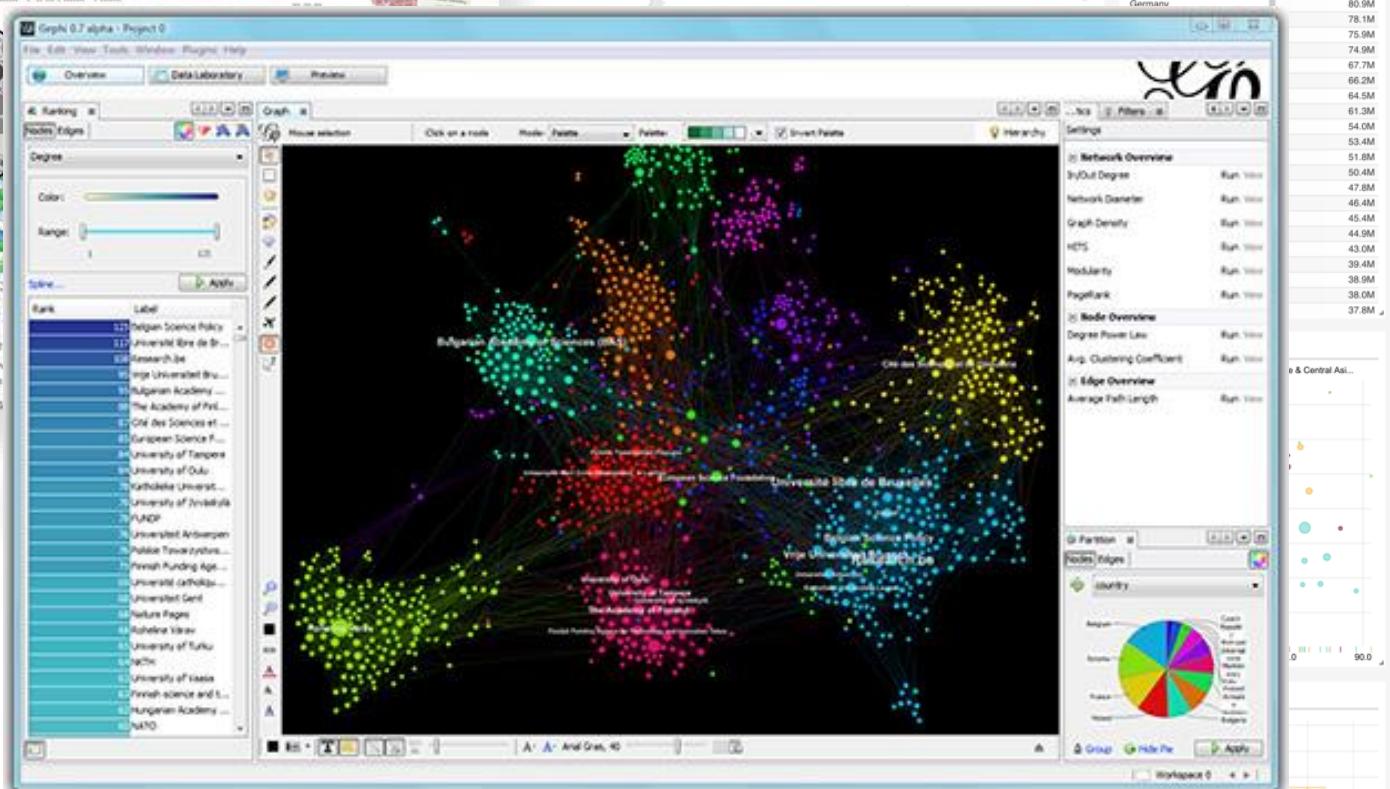
# Apache Superset



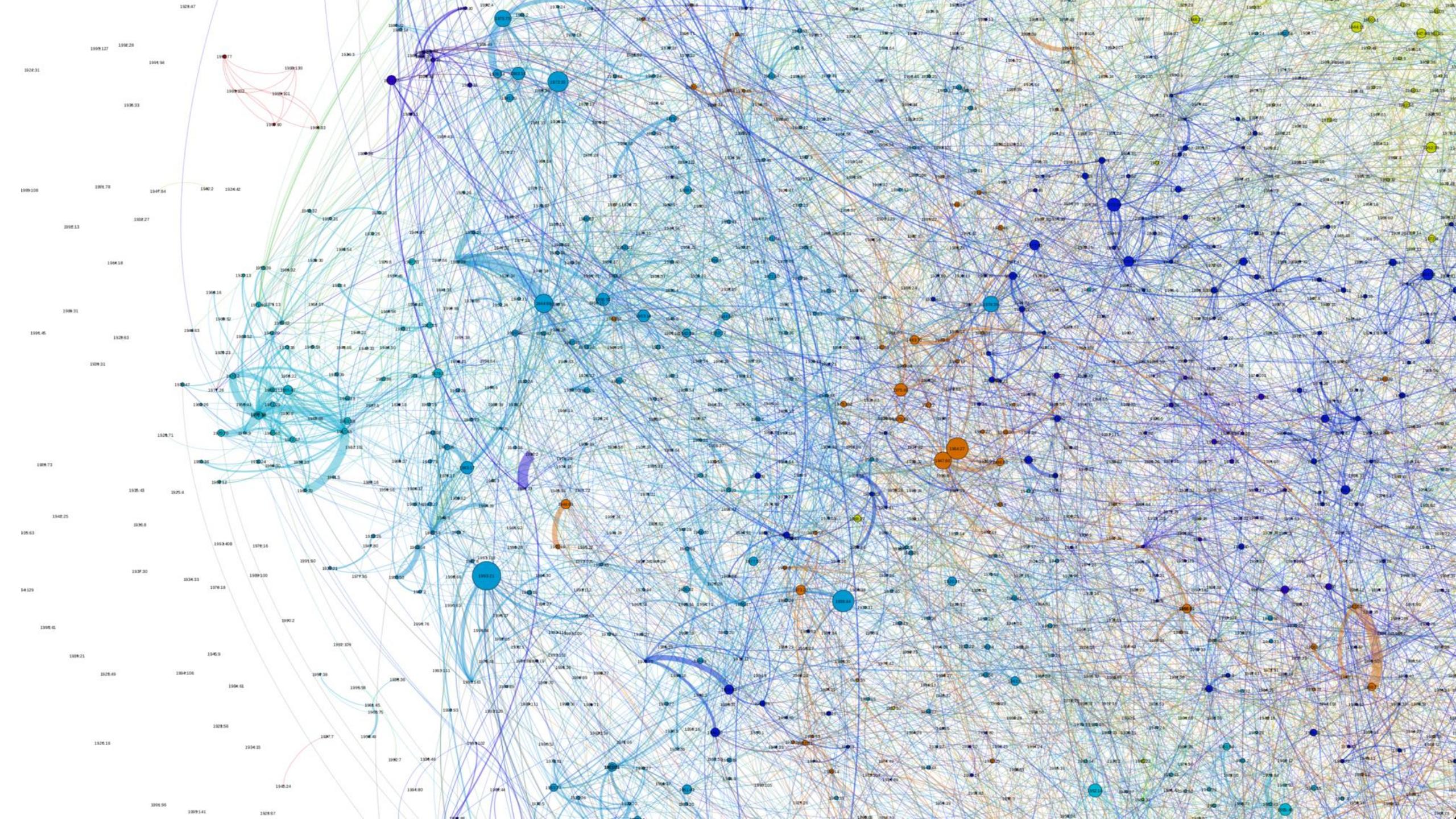
## Tableau Public

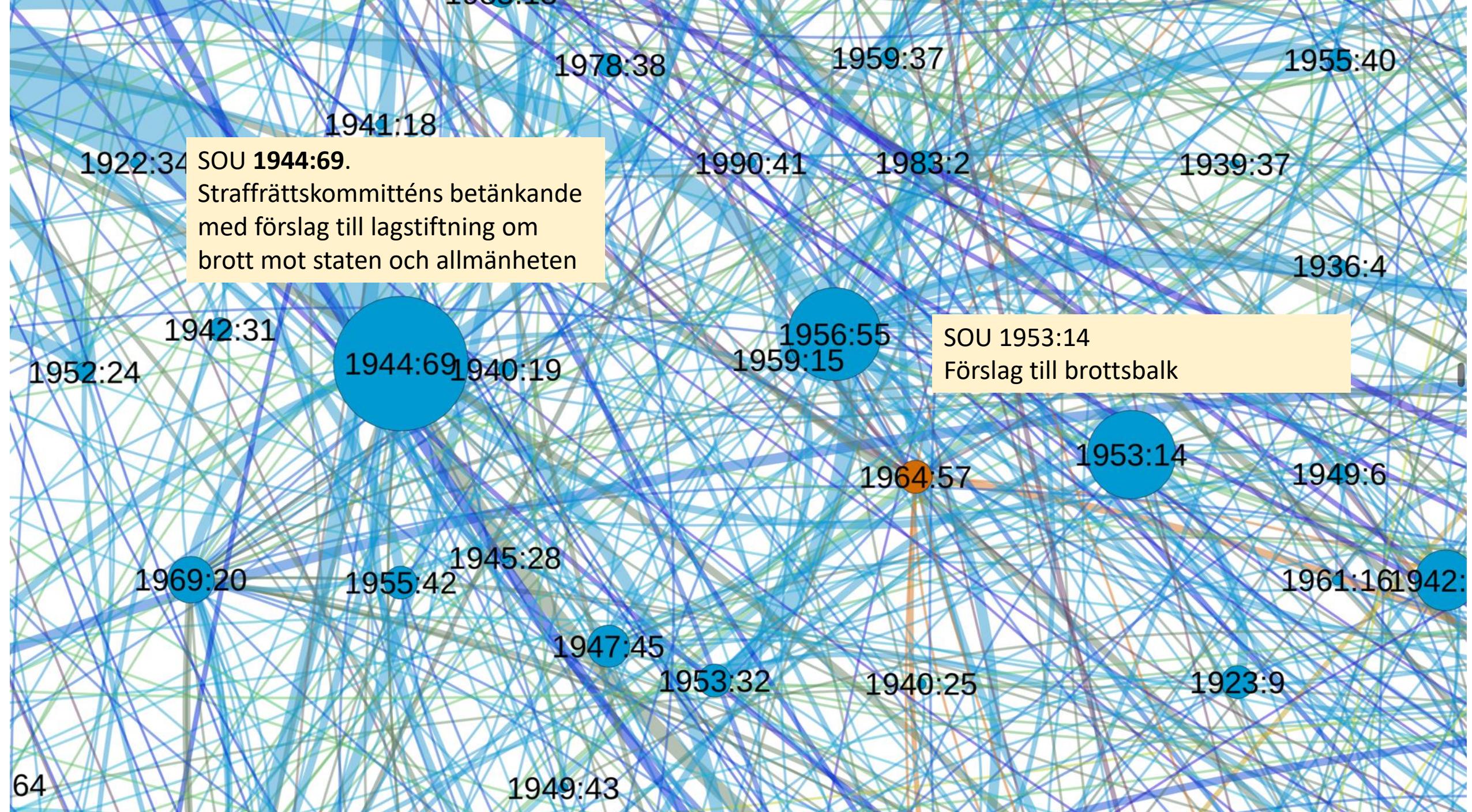


## R Statistical Computing



Statens offentliga utredningar  
(självciteringar) + Gephi





Eurostat + Minecraft  
(offentliga utgifter enligt COFOG)



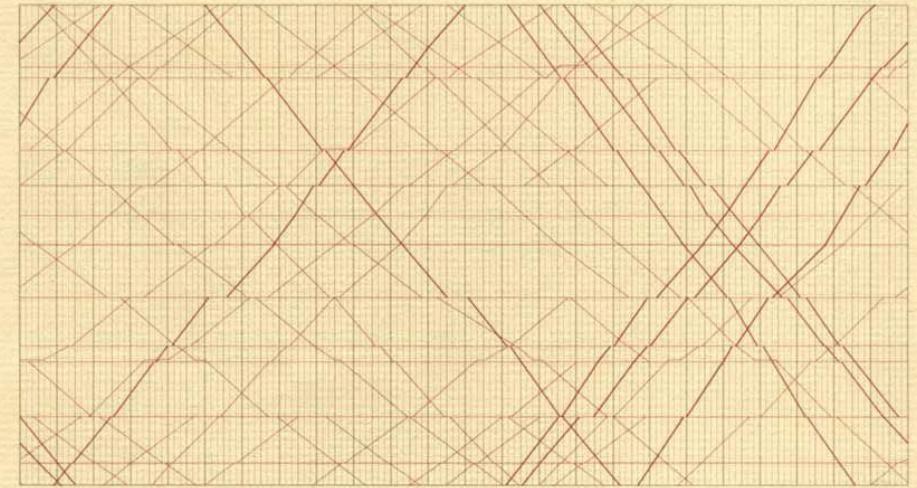


...men blir det begripligt?

Leder det till att någon tar bättre beslut?

# The Visual Display of Quantitative Information

(Edward Tufte)



The Visual Display  
of Quantitative Information

EDWARD R. TUFTE

# Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite  
Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mohilow et se rejoignaient vers Orscha en Wilebsk, avaient toujours marché avec l'armée.

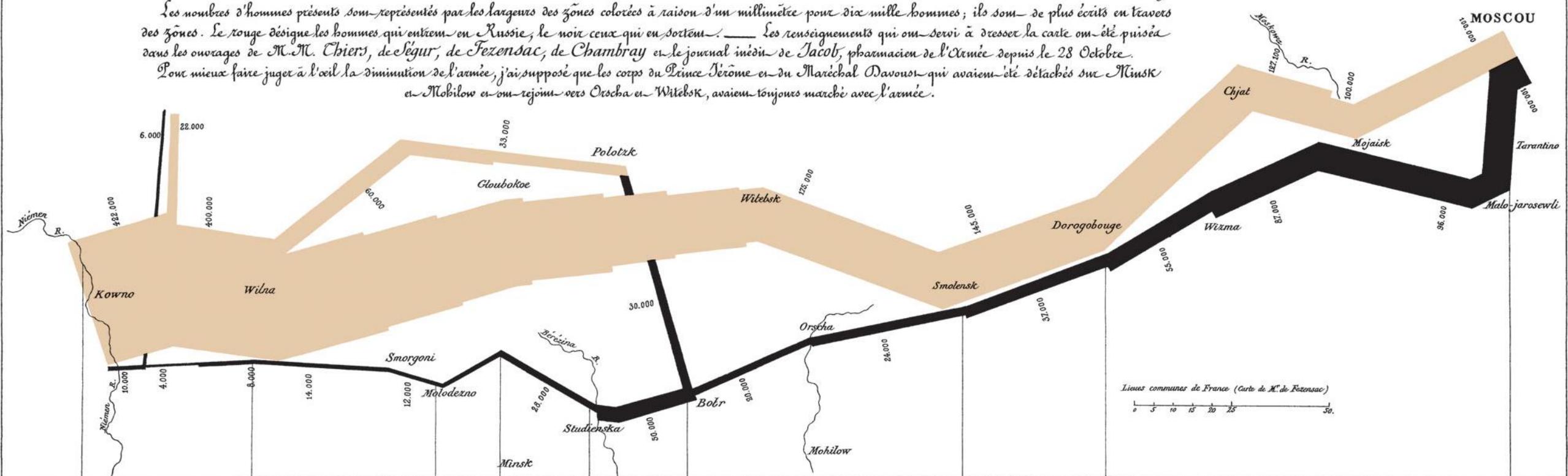
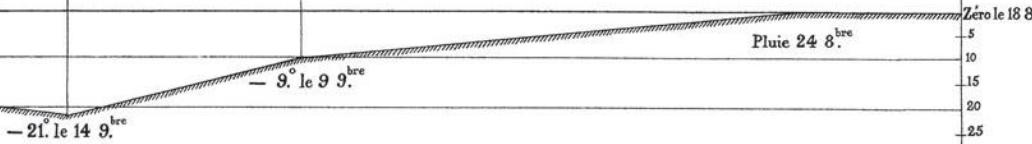


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop  
le Niemen gelé.

- 26° le 7 X.<sup>bre</sup>  
- 30° le 6 X.<sup>bre</sup>  
24. le 1<sup>er</sup> X.<sup>bre</sup>  
- 20° le 28 9.<sup>bre</sup>  
- 11°



# The Cognitive Style of PowerPoint

(Edward Tufte)



On this single Columbia slide, in a PowerPoint festival of bureaucratic hyper-rationalism, 6 different levels of hierarchy are used to classify, prioritize, and display 11 simple sentences:

|         |                                |
|---------|--------------------------------|
| Level 1 | Title of Slide                 |
| Level 2 | • Very Big Bullet              |
| Level 3 | — dash                         |
| Level 4 | * diamond                      |
| Level 5 | • little bullet                |
| Level 6 | ( ) parentheses ending level 5 |

The analysis begins with the dreaded "Executive Summary." A conclusion is presented as a headline title: "Test Data Indicates Conservatism for Tile Penetration." This turns out to be unmerited reassurance. Executives, at least those who don't want to get fooled, had better read far beyond the title.

The "conservatism" is *not* about the predicted tile damage but rather about the *choice of models* that might be used to predict damage! But why, after 112 nights, are models being calibrated during a crisis? How can "conservatism" be inferred from a loose comparison of a computer model and some thin data? Divergent evidence means divergent evidence, not inferential security. Claims of analytic "conservatism" should be viewed with skepticism. Such claims are sometimes a rhetorical tactic that substitutes verbal fudge factors for quantitative assessments.

As the analysis continues, the seemingly reassuring conclusion of the headline fades away.

These lower-level bullets at the end of the slide reveal that the headline conclusion is irrelevant and diverting. This third-level point notes that "Flight condition [that is, the Columbia] is significantly outside of test database." How far outside? The final bullet will tell us.

This fourth-level bullet concluding the slide says that, by the way, the debris that struck the Columbia is estimated to be 1920/3 = 640 times larger than data used in the tests of the model! Thus a better headline would be "Review of Test Data Indicates Irrelevance of Two Models." There is an interesting dynamic to this slide: the headline is an exercise in misdirection, which the text then awkwardly and slowly eviscerates.

The Very-Big-Bullet sentence does not seem to make sense.

Spray On Foam Insulation

A model to estimate damage to the tiles protecting the left wing

## Review of Test Data Indicates Conservatism for Tile Penetration

- The existing SOFI on tile test data used to create Crater was reviewed along with STS-73 Southwest Research data
  - Crater overpredicted penetration of tile coating significantly
    - Initial penetration is described by normal velocity
      - Varies with volume/mass of projectile (e.g., 200ft/sec for 3cu. In)
    - Significant energy is required for the softer SOFI particle to penetrate the relatively hard tile coating
      - Test results do show that it is possible at sufficient mass and velocity
    - Conversely, once tile is penetrated SOFI can cause significant damage
      - Minor variations in total energy (above penetration level) can cause significant tile damage
  - Flight condition is significantly outside of test database
    - Volume of ramp is 1920cu in vs 3 cu in for test

6

A reference to a foam insulation piece that separated from the bipod ramp tying the orbiter to the large liquid fuel tank. Instead of "ramp," say "estimated volume of one of several pieces of debris that might have damaged the wing."

\*In their final report (p. 191), the Columbia Accident Investigation Board developed this point about units of measurement: "While such inconsistencies might seem minor, in highly technical fields like aerospace engineering a misplaced decimal point or mistaken unit of measurement can easily engender inconsistencies and inaccuracies."

The vaguely quantitative words "significant" and "significantly" are used 5 times on this slide, with *defacto* meanings ranging from "detectable in largely irrelevant calibration case study" to "an amount of damage so that everyone dies" to "a difference of 640-fold." None of these 5 usages appears to refer to the technical meaning of "statistical significance."

The low resolution of PowerPoint slides promotes the use of compressed phrases like "Tile Penetration." As is the case here, such phrases may well be ambiguous. The low resolution and large font generate 3 typographic orphans, lonely words dangling on a separate line:  
Penetration significantly 3cu. in

This vague pronoun reference "it" alludes to *damage to the left wing*, which caused the destruction of the Columbia. The slide weakens important material with ambiguous language (sentence fragments, passive voice, multiple meanings of "significant"). The 3 reports were created by engineers for high-level NASA officials who were deciding whether the threat of wing damage required further investigation before the Columbia attempted to return. Satisfied that the reports indicated that the Columbia was not in danger, the officials made no further attempts to assess the threat. The slides were part of an oral presentation, later circulated as e-mail attachments.

In this slide the same unit of measure for volume (cubic inches) is shown a different way every time  
3cu. in      1920cu in      3 cu in  
rather than in clear and tidy exponential form  $1920 \text{ in}^3$ . Perhaps the available font cannot show exponents. Shakiness in conventions for units of measurement should provoke concern.\* Slides with hierarchical bullet-outlines do not handle statistical data and scientific notation gracefully. If PowerPoint is a corporate-mandated format for all engineering reports, then some competent scientific typography (rather than the PP market-pitch style) is essential. In this slide, the typography is so choppy and clunky that it impedes understanding.

# Avslutningsvis: 7 medskick

- Utforska sätt för visualisering av information (men utan att tappa målet om vad du vill berätta)
- Testa visualiseringar med användare
- Gör information möjlig att länka ihop med annan information
- Maximera upplösning
- Öppna upp data (inklusive kodverk i maskinläsbar form)
- Automatisera insamlingen av standardiserad information direkt från källsystem
- Lär dig grunderna i programmering

Nyttan med information  
uppstår när den begrips

Tack!

peter.krantz@skl.se