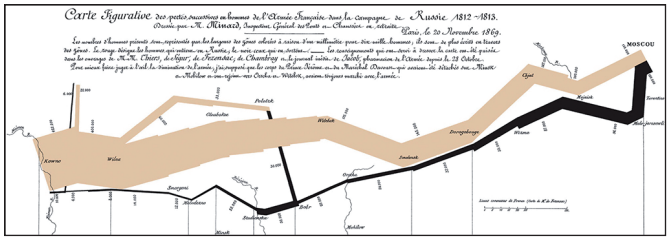
Pauline Vos, Eric Oziegbe Omobude

# Sankey-diagrammen in de wiskundeles

Pauline en Eric werken in Noorwegen, waar ze met leerlingen uit klas 2 (vmbo-havo-vwo) bij een bedrijf op bezoek gaan. Terug in de klas maken leerlingen wiskundeopdrachten die bij de context van het bedrijf aansluiten. Het ontwerpen van die opdrachten is lastig, omdat ze ook moeten passen bij de interesse en competenties van de leerlingen. Sankey-diagramen bieden uitkomst, omdat je er *holistisch* mee naar bedrijfsprocessen kunt kijken.

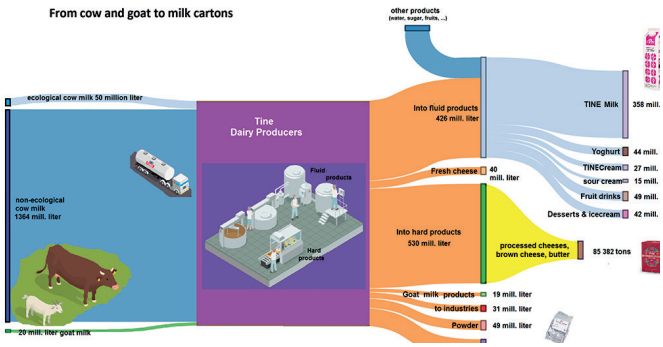
## Wat zijn Sankey-diagrammen?

Sankey-diagrammen zijn stroomdiagrammen waarbij de breedte van de stroom proportioneel is met de hoeveelheid. De lengte doet er niet toe, en deze flexibiliteit maakt dat er creatief mee gevisualiseerd kan worden. Een van de eerste Sankey-diagrammen werd in 1869 gemaakt door Charles Minard over de Russische veldtocht van Napoleon, zie figuur 1.




*figuur 1* [1]

Een brede, beige stroom toont hoe 422.000 Franse soldaten in de herfst op weg naar Moskou gingen. Vervolgens wordt de stroom soldaten dunner. Een kwart kwam aan in Moskou, op de terugweg wordt de zwarte stroom nog dunner. Een pieterig stroompje geeft weer dat 2% van alle soldaten uit die oorlogswinter terugkeerde. Minard gebruikte handig de landkaart als achtergrond om de stroom soldaten, tijdstip, plaats, temperatuur, de richting van de veldtocht en de dramatische sterfte onder de soldaten overzichtelijk weer te geven.
Pas later werd er een naam gegeven aan dit type diagram. Het is een beetje ongelukkig dat de naamgever van de Sankey-diagrammen, Capt Matthew Sankey, niet de eerste was die ze maakte. Hij maakte rond 1900 diagrammen om energiestromen in stoommachines weer te geven, namelijk hoeveel energie er vanuit een boiler wordt omgezet in arbeid, en hoeveel er verloren gaat aan wrijving, warmtestraling en rook. Sankey-diagrammen kun je dus gebruiken om de stroom van mensen, energie, geld of andere zaken weer te geven, en zijn daarmee geschikt voor het visualiseren en modelleren van complexe processen zoals te vinden in de bedrijven die onze leerlingen bezoeken. In figuur 2 staat een Sankey-diagram van de liters melk die van de koeien bij de boer naar de Noorse melkfabriek Tine AS gaan, daar omgezet worden in vloeibare en harde melkproducten, die vervolgens naar de consument komen als yoghurt, ijs, kaas en andere zuivelproducten. Dit diagram is gebaseerd op authentieke data uit het jaarverslag van deze Noorse melkfabrikant en gebruikten we in onze lessen om Sankey-diagrammen te introduceren[2].

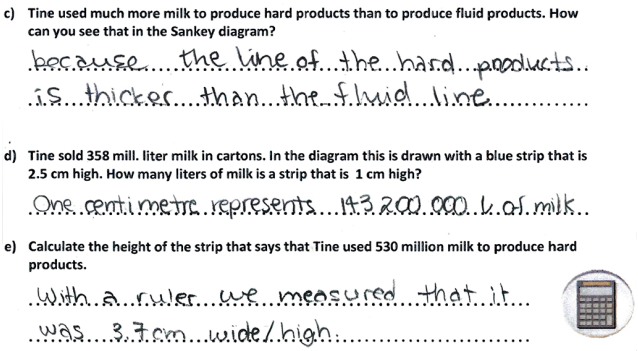



*figuur 2* [2]

## Lesmateriaal en ervaringen uit de les

Pauline had een werkboekje van acht pagina's ontworpen[3] en in kleine kring uitgetest. De opbouw van het boekje is dat de leerlingen eerst enkele diagrammen moeten interpreteren, en daarna zelf Sankey-diagrammen gaan tekenen. Eric heeft het materiaal op zijn school in een

tweede klas (*grade 8*) met negentien leerlingen uitgeprobeerd, waarbij de leerlingen in groepjes van drie of vier werkten. De leerlingen hadden twee lesuren nodig om het boekje door te werken, maar door corona zat er helaas meer dan een maand tussen die twee lessen. Toen de scholen weer opengingen en Eric zijn klas vroeg of ze het nog wilden afmaken, was het antwoord unaniem positief. Figuur 3 geeft een indruk van enkele opdrachten. Deze gaan over het interpreteren van de stroken als hoeveelheden melk, en het berekenen van de breedte van de stroken. Uit de antwoorden van de leerlingen konden we opmaken dat ze het idee van een Sankey-diagram goed oppikten, maar ook dat ze liever de breedte met een liniaal opmaten dan die te berekenen.


c) Tine used much more milk to produce hard products than to produce fluid products. How can you see that in the Sankey diagram?

because the line of the hard products is thicker than the fluid line

d) Tine sold 358 mill. liter milk in cartons. In the diagram this is drawn with a blue strip that is 2.5 cm high. How many liters of milk is a strip that is 1 cm high?

One centimetre represents 143 200 000 L of milk

e) Calculate the height of the strip that says that Tine used 530 million milk to produce hard products.

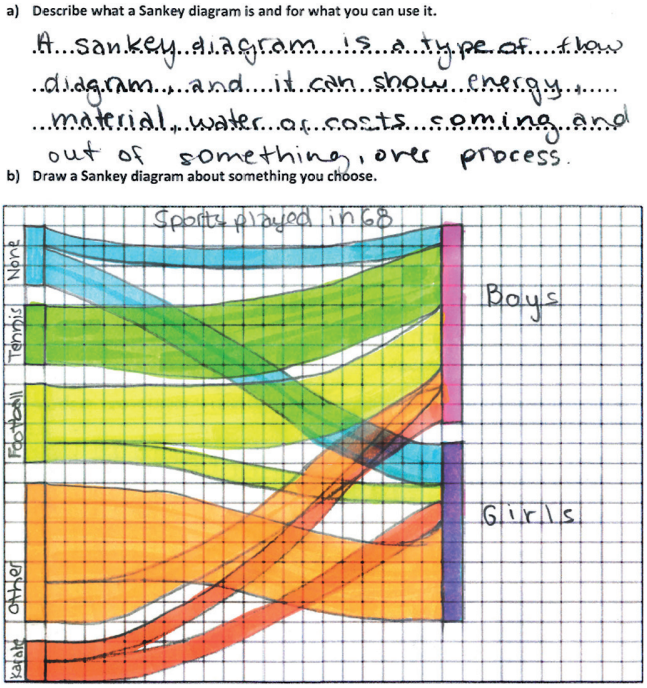
With a ruler we measured that it was 3.7 cm wide/high

*figuur 3 Opdrachten uit het boekje en antwoorden van leerlingen over de breedte van de stroken in het Sankey-diagram van de melkfabriek*

Op de laatste pagina van het boekje vroegen we de leerlingen om een definitie van Sankey-diagrammen te geven en om er een te tekenen over een onderwerp naar keuze. Figuur 4 geeft de antwoorden van een groepje weer. De leerlingen geven als definitie vooral aan dat je met Sankey-diagrammen processen kunt weergeven. Dit was wat alle groepjes schreven: ze zagen vooral nut in de Sankey-diagrammen als input-outputdiagram en geen enkel leerlingengroepje meldde iets over de breedte van de stroken, hoewel ze hier wel opgaven over gemaakt hadden (zoals in figuur 2). Blijkbaar springt het kwalitatieve aspect van de diagrammen (dat je er processen mee kunt aangeven) er meer uit voor de leerlingen dan het kwantitatieve (dat je er hoeveelheden mee kunt aangeven)

## Data met verschillende categorieën

Op de vraag om een Sankey-diagram naar keuze te maken zagen we vaak de hobby's van de leerlingen terug. In de meeste gevallen hadden ze dan evengoed een cirkeldiagram kunnen maken. In figuur 4 zien we een uitzondering hierop. Dit groepje had hobby's en sekse aan


a) Describe what a Sankey diagram is and for what you can use it.

A sankey diagram is a type of flow diagram and it can show energy material water or costs coming and out of something, over process.

b) Draw a Sankey diagram about something you choose.



*figuur 4 Antwoorden van een leerlingengroepje over een definitie en een Sankey-diagram naar keuze*

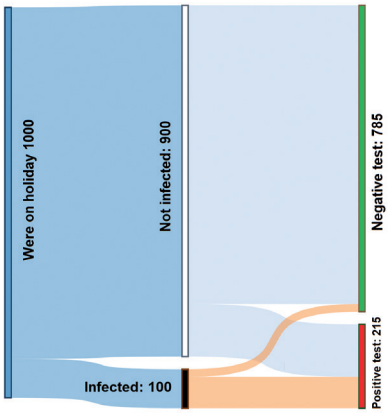
elkaar gekoppeld binnen een Sankey-diagram. Indien een Sankey-diagram op deze manier wordt gebruikt, dan wordt deze ook wel aangeduid als *alluval plot* of *parallel sets*. Op deze manier kunnen relaties tussen *categorieën* binnen een dataset gevisualiseerd worden. Zo kun je op het internet Sankey-diagrammen vinden over de passagiers op de eerste (en laatste) reis van de Titanic, waarbij elke passagier gelabeld is volgens de vier categorieën: passagiersklasse, geslacht, volwassene/kind en wel/niet overleefd.[4] Met dit Sankey-diagram tonen de breedte van de stroken duidelijk aan dat vooral mannen uit de bemanning en de laagste passagiersklasse overleden. Een ander voorbeeld is een sociologisch overzicht van moordzaken in de VS, waarbij de categorieën wapen, etniciteit en sekse van slachtoffer en dader, hun relatie en aanleiding voor de moord.[5] Dit diagram laat door de breedte van de stroken duidelijk zien dat de meeste slachtoffers man zijn, evenals de daders, dat ze elkaar meestal kenden en dat de aanleiding meestal een ruzie was. Door de breedte van de stroken te vergelijken kun je ook zien dat niet-witte mannen vaker door een witte man worden vermoord dan andersom. Deze twee voorbeelden hadden we niet opgenomen in het lesmateriaal, maar geven aan dat Sankey-diagrammen mogelijkheden bieden voor interdisciplinaire lessen op het snijvlak van wiskunde en de maatschappijvakken.

>

## Fout-positief en fout-negatief

We hadden nog een opgave ontworpen die buiten het boekje is gebleven, maar die we je niet willen onthouden. De opgave gaat over fout-positieve en fout-negatieve medische testresultaten.[6] De opdracht is actueel, hoewel deze werd ontworpen voordat de coronaepidemie toesloeg. De opdracht gaat over een fictieve groep van duizend vakantiegangers, die op hun terugreis horen dat 10% van hen besmet is met een virus en dat het een tijd duurt voordat de ziektesymptomen komen. Alle reizigers moeten daarom direct een test doen. Deze test is alleen niet heel precies: 80% van de besmette reizigers krijgt een positief resultaat en 85% van de niet-besmette reizigers krijgt een negatief resultaat. In figuur 5 zijn de duizend reizigers in een Sankey-diagram gevisualiseerd, uitgesplitst naar besmetting en testresultaat. Je kunt de leerlingen allerlei vragen stellen, onder andere: moet een reiziger met een positieve uitslag zich zorgen maken? Het diagram laat zien dat als er weinig besmette mensen in een populatie zijn, het aantal fout-positieve uitslagen relatief hoog is. Dit komt doordat een klein deel van een grote groep meer kan zijn dan een groot deel van een kleine groep (lees deze zin rustig nog eens). Het diagram geeft aanleiding tot wiskundig redeneren met bijvoorbeeld enkele vervolgvragen van het type 'what if..?':

— Hoe verandert het aantal fout-positieven als de besmettingsgraad hoger wordt?
— Voorspel hoe het Sankey-diagram eruitziet als je het uitbreidt met een derde categorie: de resultaten van de her-test.




*figuur 5 Fout-positieve, negatieve en fout-negatieve testresultaten*[2]

## Ten slotte

We denken dat Sankey-diagrammen veel mogelijkheden bieden voor interessante wiskundelessen. Enerzijds nodigen ze uit tot wiskundig redeneren en gaan ze over mooie maatschappelijk relevante contexten. Anderzijds zijn ze goed bruikbaar op allerlei niveaus, van onderbouw, in het vmbo, tot in de bovenbouw bijvoorbeeld in de lessen over hypothesetoesten bij statistiek. De veelzijdigheid van Sankey-diagrammen komt doordat ze twee rollen tegelijk spelen. Enerzijds kun je met Sankey-diagrammen complexe processen *modelleren*. Als model horen ze dan altijd bij een fenomeen; ze zijn een model van *iets*. Ze zijn dan een tool om de werkelijkheid wiskundig te vatten. Anderzijds behoren Sankey-diagrammen tot de *beeldtaal*; ze zijn familie van (data-)visualisaties, zoals ook histogrammen, boxplots, treemaps, Ganttmaps, enzovoort. Ze zijn daarmee een tool om te communiceren. Daardoor zien we ze steeds vaker in allerlei media en in wetenschappelijke artikelen. Met de groeiende hoeveelheid data in onze maatschappij zien we ook een groeiende hoeveelheid visueel-wiskundige tools die helpen om overzicht in die Big Data te scheppen. Als wiskundedocenten kunnen we leerlingen helpen om deze beeldtaal beter te interpreteren, en om deze zelf te leren hanteren.[7]
Ons lesmateriaal is vrij beschikbaar. We werken aan vervolgmateriaal, onder andere aan Sankey-diagrammen over de processen in de bedrijven in onze regio's (onder andere het olie-, de vuilverbrandings- en het riolerings-bedrijf). Helaas staan onze bedrijfsexcursies met de klassen momenteel 'on hold', omdat bedrijven huiverig zijn voor bezoek. Daarom denken we verder over meer digitaal materiaal. Indien je mee wilt denken, het materiaal naar het Nederlands wilt vertalen, het materiaal wilt aanpassen, of iets anders met Sankey-diagrammen wilt, dan horen we het graag.

Het lesmateriaal over Sankey-diagrammen vind je op

*vakbladeuclides.nl/967sankey*

## Noten

[1] Zie https://upload.wikimedia.org/wikipedia/commons/2/29/Minard.png
[2] Een beter leesbare versie van deze figuur is te zien in [3].
[3] Werkblad *Sankey diagrams* (in het Engels), zie *https://www.researchgate.net/publication/339460457_Sankey_diagrams*
[4] Zie bijvoorbeeld *https://cran.r-project.org/web/packages/alluvial/vignettes/alluvial.html*
[5] Zie: *https://medium.economist.com/an-anatomy-of-murder-20ef83346d37*

[6] Fout-negatief (vals-positief) en fout-negatief (vals-negatief) staan ook wel bekend, respectievelijk, als fout van de eerste soort (Type I fout) en fout van de tweede soort (Type II fout). In de statistiek wordt een fout van de eerste soort gemaakt als een statistische toets een nulhypothese onterecht verwerpt, en een fout van de tweede soort wordt gemaakt, als een statistische toets een nulhypothese onterecht aanneemt.

[7] Er zijn verschillende manieren om Sankey-diagrammen te maken. We raden aan dat leerlingen eerst de diagrammen met de hand maken. Onze leerlingen vonden daarna zelf een eenvoudig, gratis tooltje, zie *www.sankeymatic.com.*

## Over de auteurs

Pauline Vos is hoogleraar Mathematics Education aan de Universiteit van Agder (Noorwegen). Naast het opleiden van leraren leidt ze onderzoek, onder andere naar relevantie-ervaringen van leerlingen in het wiskundeonderwijs ('juf, waarom moet ik deze sommen doen?'). E-mailadres: *pauline.vos@uia.no*
Eric Oziegbe Omobude is wiskundeleraar aan de Arendal International School (Noorwegen) en IB-examinator. Hij deed zijn masters in Mathematics Education aan de Universiteit van Agder en is nieuwsgierig naar hoe leerlingen wiskunde leren door middel van wiskundig modelleren.
E-mailadres: *ericomobude@gmail.com*

# Sankey diagrams in the mathematics class

Pauline Vos and Eric Oziegbe Omobude share their experience of teaching Sankey diagrams.

Sankey diagrams are useful for modeling and visualizing industrial processes. They are also useful tools for visualizing relationships within complex data sets. This allows teachers to visualise how false positive and false negative test results are related.

We, Pauline and Eric, both work in Norway, where we try to connect mathematics meaningfully to enterprises in the region. However, it is hard to design meaningful mathematical assignments that match the interests and competences of the students, as well as match industrial contexts of enterprises. We tried to use industrial context for mathematical calculations, such as using oil tanks for tasks on volume calculations or using a pressure gauge for a task to translate the PSI scale into a BAR scale. However, such tasks situated on details of the industrial processes did not assist students in better understanding the industrial processes, nor how mathematics is useful at workplaces. This made us search for a mathematical tool that can capture industrial processes holistically and came up with Sankey diagrams.

### What are Sankey diagrams?

Sankey diagrams are flow charts, in which the width of a flow is proportional to the quantity represented. The three Sankey diagrams in Figure 1 show the energy conversion in different lightbulbs. The three bulbs emit the same amount of light, but the loss of electrical energy into heat differs considerably between the three types. In a Sankey diagram, the length of a flow can vary, which allows for flexibility and creativity in visualising data.

The length of a flow can vary, which allow for flexibility and creativity in visualizing data. In 1869 one of the first ever Sankey diagrams was created by Charles Minard. He visualized Napoleon's Russian campaign. The start of the black flow is wide and shows how 422,000 French soldiers went on the march to Moscow in the fall of 1812. The flow of soldiers becomes thinner because of battles, hunger, sicknesses, and drowning when crossing rivers. One out of four soldiers survived to arrive in Moscow, which is shown by the flow having a width, which is a quarter of the starting flow. Upon returning home, the black flow thins even further. A tiny flow shows that 2% of all soldiers returned from that winter war (See https://upload.wikimedia.org/wikipedia/commons/2/29/Minard.png). Minard cleverly used the geography as a background to clearly display the numbers of soldiers, time, place, temperature, direction of the campaign and the dramatic mortality among the soldiers.

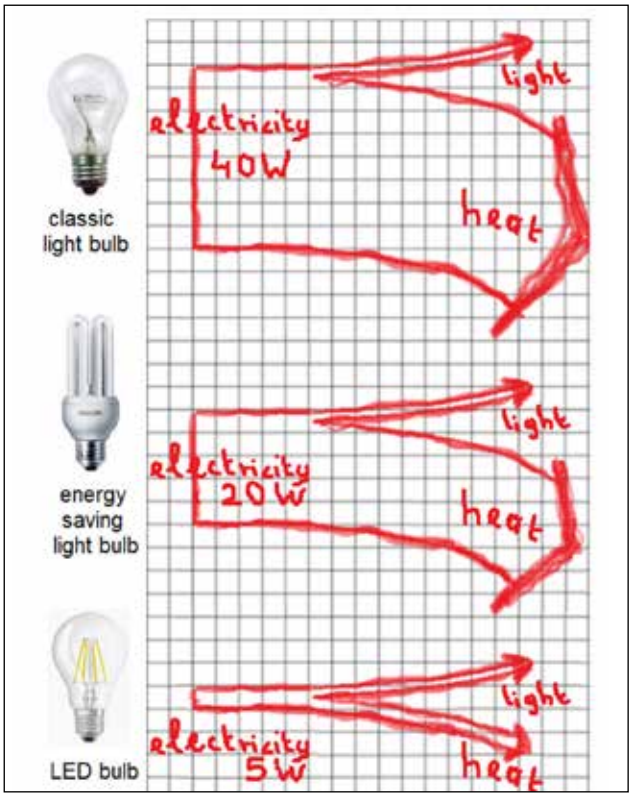



Figure 1: Sankey diagrams.

A few decades later, this type of diagram received a name. It is a bit unfortunate that the namesake of the Sankey diagrams, Captain Matthew Sankey, was not the first to make them. Around 1900 he made diagrams to show energy flows in steam engines, namely how much energy is converted from a boiler into work, and how much is lost to friction, heat radiation and smoke.

Sankey diagrams can be used to display the flow of people, energy, money or other things, and are therefore suitable for visualizing and modeling complex processes as found in the enterprises in our

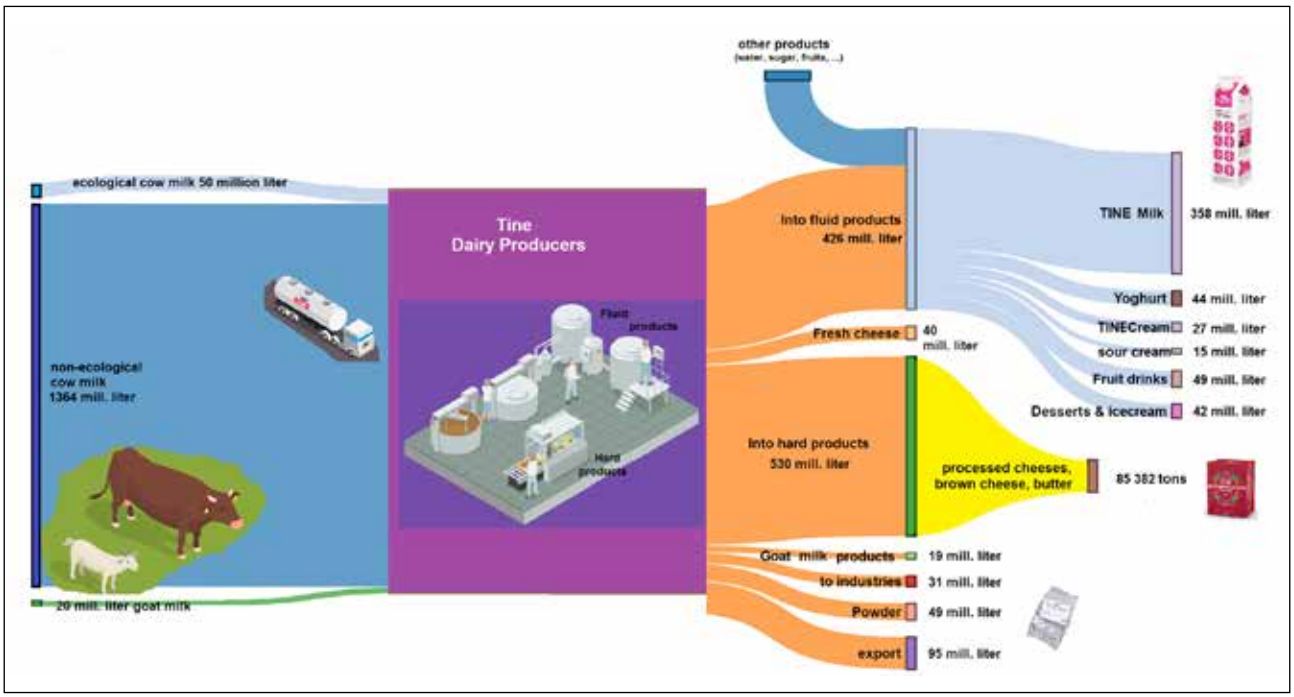


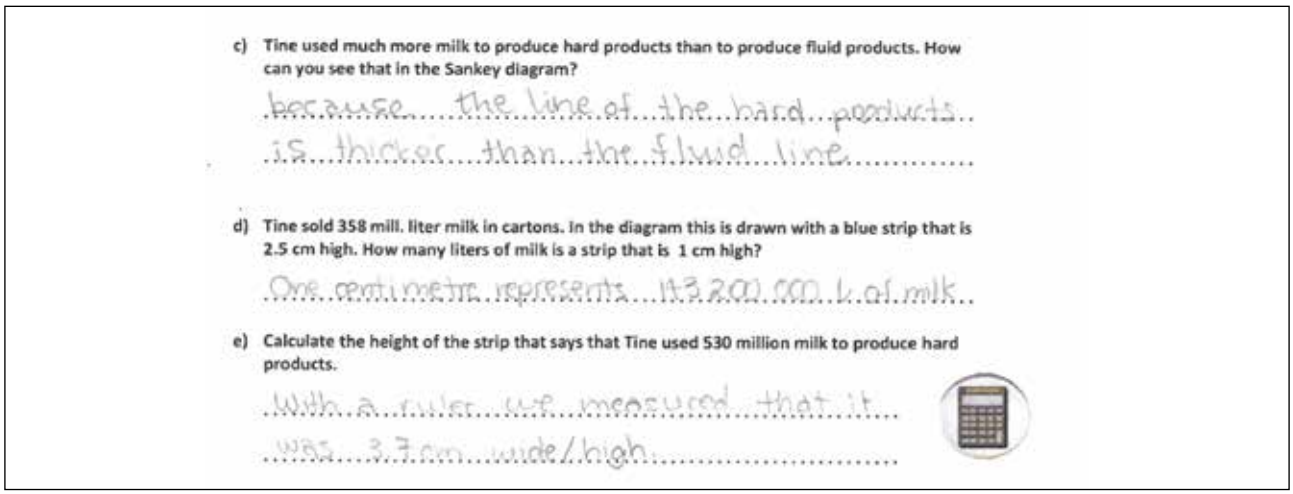

Figure 2: From cow and goat to milk cartons

region. Figure 2 shows a Sankey diagram of the litres of milk that go from the farm's cows to the Norwegian dairy factory Tine AS, where it is converted into fluid and hard milk products, which are then brought to the consumers as yoghurt, ice cream, cheese and other dairy products. This diagram is based on authentic data from the annual report of this Norwegian milk manufacturer and we used it in our lessons to introduce Sankey diagrams.

## Teaching materials and experiences from the lesson

Pauline had designed an 8-page booklet and tested it at a small scale with a few students. The structure of the booklet is that the students first have to interpret diagrams, and then draw Sankey diagrams themselves. Eric tested the material at his school with 19 14-year-old students, working in groups of three or four. The students needed two lessons to work through the booklet, but due to the school closure due to the corona pandemic, unfortunately there was more than a month between those two lessons. When the schools reopened and Eric asked his class if they wanted to finish it, the answer was unanimously positive.

Figure 3 shows some of the tasks. These are about interpreting the flows as quantities of milk and calculating the width of the flows. From the students'



c) Tine used much more milk to produce hard products than to produce fluid products. How can you see that in the Sankey diagram?

because the line of the hard products is thicker than the fluid line

d) Tine sold 358 mill. liter milk in cartons. In the diagram this is drawn with a blue strip that is 2.5 cm high. How many liters of milk is a strip that is 1 cm high?

One centimetre represents 143 200 000 L of milk

e) Calculate the height of the strip that says that Tine used 530 million milk to produce hard products.

With a ruler we measured that it was 3.7 cm wide/high

Figure 3: Tasks from the booklet and students' answers about the widths of the flows in the Sankey diagram of the dairy enterprise.

answers, we could see that they appropriated well the idea of a Sankey diagram, but also that they would rather measure the width with a ruler than calculate it.

### Teaching materials and experiences from the lesson

On the last page of the booklet we asked the students to define and draw Sankey diagrams on a topic of their choice. Figure 4 shows the answers of one group in Eric's class. As a definition, the students mainly indicated that one can use Sankey diagrams to represent processes. This was what all groups wrote: they saw the Sankey diagrams especially useful as an input-output diagram; not one student group reported anything about the width of the flows, although they had made calculations about this (as in Figure 2). Apparently, the qualitative property of the diagrams (that one can indicate processes with them) stands out more for the students than the quantitative property (that one can use them to indicate quantities and relations between these).

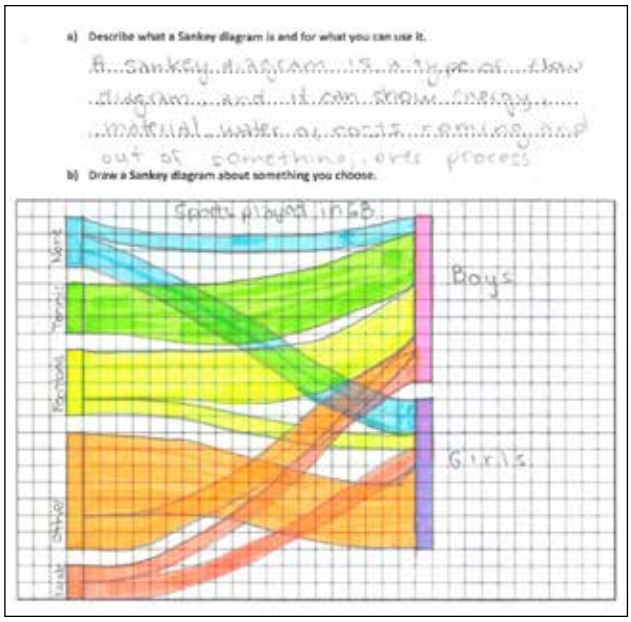



Figure 4: Answers of a student group about the definition and a free Sankey diagram.

### Sankey diagrams for data with different categories

When asked to make a Sankey diagram themselves, we often saw the hobbies of the students. In most cases this type of information could just as well be visualized through a pie chart. In Figure 4 we see an exception to this. This group had linked hobbies and gender within the Sankey diagram. When a Sankey diagram is used in this way, it is also referred to as *alluval plot* or *parallel sets*. In this way, relationships between categories within a dataset can be visualized.

For example, on the internet, one can find Sankey diagrams about the passengers on the first (and last) voyage of the Titanic, with each passenger labelled according to the four categories of (1) passenger class, (2) gender, (3) adult / child and (4) survived / not. With such a Sankey diagram, the width of the flows clearly displays that mainly men from the crew and the lowest passenger class perished. Another example is a sociological overview of murder cases in the US, in which the categories weapon, ethnicity and sex of victim and perpetrator, their relationship and cause for the murder. The width of the flows in this diagram clearly show that most of the victims are men, as are the perpetrators, that they usually knew each other and that the cause was usually an argument. By comparing the width of the flows, one can also see that non-white men are more often killed by a white men than the other way around. We had not included these examples (passengers on the Titanic and murder cases in the US) in the teaching material. However, these examples show that Sankey diagrams offer opportunities for interdisciplinary lessons at the intersection of mathematics and social subjects.

### False positive and false negative

We had designed another task, which remained outside the booklet, but which we would like to share with teachers of statistics, in particular regarding hypothesis testing. The task concerns false positive and false negative medical test results. False negative errors and false positive errors are also known as Type I and Type II error, respectively. In statistics, a Type I error occurs when a statistical test incorrectly rejects the null hypothesis, and a Type II error occurs when a statistical test incorrectly accepts the null hypothesis. The assignment is contemporary, although it was designed before the corona epidemic hit. The task is about a fictional group of 1000 holidaymakers, who on their return journey learn that 10% of them are infected with a virus and that it takes a while for the sickness symptoms to appear. All travellers must therefore immediately take a test. However, this test is not very precise: 80% of the infected get a positive result and 85% of the non-infected get a negative result. Figure 5 shows the 1000 travellers in a Sankey diagram, broken down by infection and test result.

This Sankey diagram allows a teacher to ask many questions, among which: should a traveller with a positive result be concerned? The diagram shows that when there are few infected people in a population, the number of false positive results is relatively high.

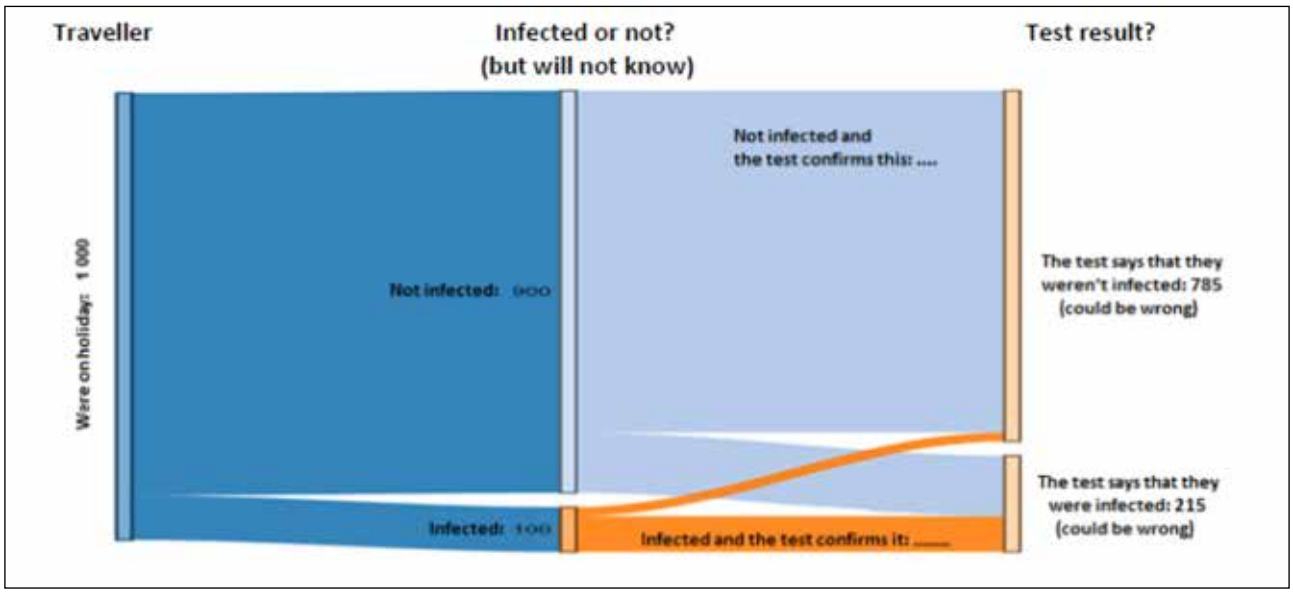



Figure 5: Sankey diagram for a medical test with false positive and false negative test results.

This is because a small part of a large group can be more than a large part of a small group (if unclear, just read the sentence again). The diagram gives rise to mathematical reasoning with for example some follow-up questions of the type "what if … ?":

- How does the number of false positives change as the infection rate increases?
- Predict what the Sankey diagram will look like if you expand with a third category: the results of the re-test.

## Last but not least

We think Sankey diagrams offer many opportunities for interesting mathematics lessons. On the one hand, they invite for mathematical reasoning and they can deal with contexts that are socially relevant to students. On the other hand, they are very useful at all levels, from lower secondary school, in pre-vocational secondary education to upper secondary education, for example in the lessons on hypothesis testing in statistics. The versatility of Sankey diagrams is due to the fact that they play two roles at the same time. On the one hand, one can model complex processes with Sankey diagrams. As a model they always belong to a phenomenon; they are a model of something. They are then a tool for understanding reality through mathematical means. On the other hand, Sankey diagrams belong to the visual language; they are a family of (data) visualisations, such as histograms, box plots, treemaps, Ganttmaps, and so forth. Therefore, they are tools in communication. As a result, we see them more and more in all kinds of media and in scientific articles. With the growing amount of data in our society, we also see a growing amount of visual-mathematical tools that help to create an overview in these Big Data. As mathematics teachers, we can help students to better interpret this visual language, and to learn to use it themselves.

Our teaching materials are freely available (see notes). We are working on follow-up material, including Sankey diagrams about the processes of more enterprises in our regions (including the oil company, the waste incinerator and the sewerage company). If you want to contribute ideas, translate the material into other languages, adapt the material, or want something else with Sankey diagrams, please let us know.

## Notes

There are several ways to create Sankey diagrams. We recommend that students create the diagrams by hand first. Our students then found a simple, free tool themselves, see www.sankeymatic.com .

Our teaching materials can be found at https://www.researchgate.net/publication/339460457_Sankey_diagrams

**Pauline Vos is Professor of Mathematics Education at the University of Agder (Norway). In addition to training teachers, she leads research, among other things into relevance experiences of students in mathematics education ("teacher, why do I have to do these sums?").**

**Eric Oziegbe Omobude is a mathematics teacher at Arendal International School (Norway) and IB examiner. He did his masters in Mathematics Education at the University of Agder and is curious about how students learn mathematics through mathematical modeling.**