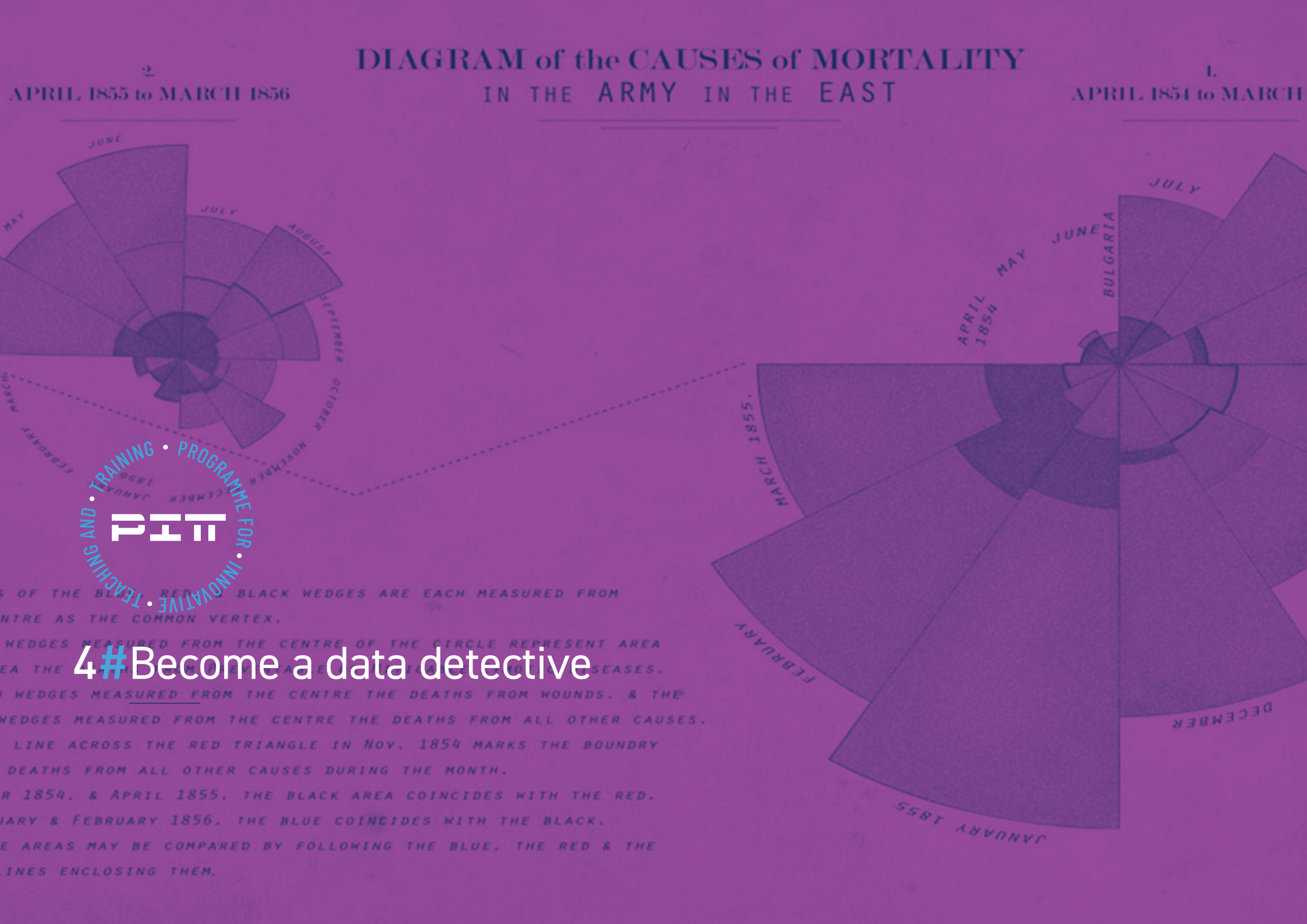


TEACHING AND TRAINING • PROGRAMME FOR INNOVATIVE •

PIT

4#Become a data detective

4.1 Didactic commentary

Ann Kiefer & Khoi Mai Huy

These days, it is essential to understand the issues at stake in debates on education, society, politics, the environment and health. To do this, it is essential to master the interpretation of statistical data and make informed judgements about the validity of the results. Statistics are ubiquitous in modern society, and understanding them is an essential part of making informed decisions.

Many authors stress that current and future citizens must be able to understand and interpret statistics to make informed decisions, especially given the abundance of data in the media (Cobb, 1999; Lajoie, Jacob and Lavigne, 1995). There is a risk that the general public will readily accept this data without critically evaluating its validity. Moreover, data are frequently selected or presented in ways intended to influence public opinion or behavior - whether by political leaders, the media, or advertising aimed at the general public. To counter these dangers of statistical manipulation by the media and politicians, Huff (1954), in his book *How to lie with statistics*, was already emphasising the need for a statistical language at the time of the Cold War, because otherwise all citizens run the risk of statistical ignorance, which means that they will be more susceptible to manipulation. More recently, Brest (2012) re-emphasises this omnipresent danger in his famous book *Damned lies and Statistics*. In it, the author presents and discusses the best examples from recent political debates that illustrate the challenges of improving statistical literacy.

Teaching statistics at school is crucial to creating informed citizens capable of analysing quantitative information. Historically, statistics education has focused on computational skills and statistical calculation procedures. However, more recent approaches to teaching and learning statistics, starting at primary level in some Western countries such as Canada (Ministère de l'éducation du Québec, 2001), place greater emphasis on the development of statistical thinking, understanding of concepts and the ability to interpret results. In Luxembourg, these skills are included in the mathematics curricula for the lower cycles of general and lower secondary education[1] . One of the areas of competence in the *Reference Guide for Media Literacy*[2] also concerns information and data.

A number of studies have highlighted the importance of thinking critically about statistical data (Shaughnessy, 2007; Shaughnessy, Garfield and Greer, 1996). Several authors (Cobb, 1999; Gattuso, 2011; Lajoie, Jacob and Lavigne, 1995; McClain, Cobb and Gravemeijer, 2000; McNab et al., 2006; Whitin, 2006) argue that the process of thinking with and about statistical data must be carried out with critical thinking in mind, i.e. the person in question must have the skills and posture to approach these quantitative situations. In addition, many other authors state that it is important for young people to understand meaning rather than learning solutions in order to be able to think independently, reflectively and critically about mathematics and science (Daniel, Lafortune, Pallascio and Sykes, 1996; Daniel, 2005). It is precisely critical thinking that enables pupils to organise, select and analyse data in order to draw relevant conclusions or predictions. In addition, pupils are able, at least in part, to judge and predict the importance and scope of their conclusions and communications in quantitative situations (Daniel, 2005).

[1] https://ssl.education.lu/eSchoolBooks/Web/ES

[2] https://edumedia.lu/wp-content/uploads/2024/12/Medienkompass_FR_web.pdf

Critical thinking, a flagship element of the 4Cs model, was identified in *Guiding Principles for Learning in the Twenty-first Century*, by Hughes and Acedo (2017) as a key educational issue.

Critical thinking is essential for pupils to become autonomous, independent and open-minded individuals.

It is essential in everyday life for making informed and considered decisions. It is also essential for evaluating statistical data rigorously. In maths lessons, pupils should therefore learn to ask questions, question statistical models and consider potential biases.

This is why, in this module, pupils analyse real data, have to solve concrete problems and justify their statistical choices. The idea, of course, is not to develop general critical thinking through a single statistics module, but rather to help their pupils develop a 'critical perspective' in relation to the notion of 'justice' through the processing of data. In this way, a 'critical perspective' would be a metacognitive tool for pupils to take a critical look at and question their mathematical models to arrive at a final model that they feel is fairer, more appropriate and better argued.

A second fundamental idea of this module is to show that in statistics there is no single correct answer. Several different approaches can be justified. We can even go one step further and show pupils that different perspectives on the same theme can lead to different statistical models and different conclusions.

It should also be noted that several studies highlight the importance of thinking critically about statistical data (Shaughnessy, 2007; Shaughnessy, Garfield and Greer, 1996). However, this research does not address the question of how to think critically about statistics and understand their implications for decision making in a school context. In general, there is a lack of explicit and systematic links between critical thinking and learning statistics.

References

1. Best, J. (2012). *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists* (1st ed.). University of California Press.
2. Cobb, P. (1999). Individual and Collective Mathematical Development: The Case of Statistical Data Analysis. *Mathematical Thinking and Learning*, 1(1), 5-43.
3. Daniel, M.-F. (2005). *Pour l'apprentissage d'une pensée critique au primaire*. Québec: Les Presses de l'Université du Québec.
4. Daniel, M.-F., Lafortune, L., Pallascio, R., and Sykes, P. (1996a; reprinted 1999, 2004). *Philosopher sur les mathématiques et les sciences*. Québec: Le Loup de Gouttière.
5. Gattuso, L. (2011). L'enseignement de la statistique : où, quand, comment, pourquoi pas ? *Statistiques et Enseignement*, 2(1), 5-30 http://publications-sfds.fr/index.php/StatEns/article/view/71
6. Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton.
7. Hughes, C. and Acedo, C. (2017). Guiding principles for learning in the twenty-first century. *Educational practices series*, 28. https://unesdoc.unesco.org/ark:/48223/pf0000262678
8. Lajoie, S., Jacobs, V. and Lavigne, N. (1995). Empowering Children in the Use of Statistics. *Journal of Mathematical Behavior*, 14, 401-425.

9. McClain, K., Cobb, P. and Gravemeijer, K. (2000). Supporting Students' Ways of Reasoning about Data. In M. J. Burke & R. F. Curcio (Eds.*) Learning Mathematics for a New Century*, 2000 Yearbook (pp. 174-187). Reston: National Council of Teachers of Mathematics.
10. McNab, S., Moss, J., Woodruff, E. and Nason, R. (2006). "We were nicer, but we weren't fairer!" Mathematical modeling exploring "fairness" in data management. *Thinking and Reasoning with Data and Chance*: 68th NCTM Yearbook (pp. 171-184). Reston: National Council of Teachers of Mathematics.
11. Ministère de l'éducation du Québec (2001). Programme de formation de l'école québécoise. Éducation préscolaire et enseignement primaire. Quebec City, QC: Gouvernement du Québec.
12. Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.*) Second handbook of research on mathematics teaching and learning*: National Council of Teachers of Mathematics, (pp. 957-1009). Reston, VA: National Council of Teachers of Mathematics.
13. Shaughnessy, J. M., Garfield, J., and Greer, B. (1996). Data handling. In A. Bishop et al. (Eds.), *International handbook of mathematics education* (v.1, p. 205-237). Dordrecht, Netherlands: Kluwer.
14. Whitin, D. J. (2006). Learning to talk back to a statistic. *In Thinking and Reasoning with Data and Chance*: 68th NCTM Yearbook (pp. 31-40). Reston: National Council of Teachers of Mathematics.

4.2 Lesson planning

This module was produced in collaboration with Prof. Khoi Mai Huy from the Université du Québec en Abitibi-Témiscamingue. This collaboration was partly funded by the ERA Chair in Mathematical Statistics and Data Science - SanDAL project (https://sandal.uni.lu/) of the Mathematics Department of the University of Luxembourg.

01 Conditions of unit

Target audience:	6e
Place:	A typical classroom
Materials required:	At least one computer or tablet with internet access per group of 2 pupils.
Duration:	3-4 times 45 minutes

02 Targeted skills

SKILLS TARGETED BY THE MATHEMATICS COURSE

Contents

- Choose a chart type
- Graphical interpretation
- Create a survey
- Interpreting survey results

The student is able to

- construct a table of values, an evolving graph/diagram, a bar graph/diagram, a circular graph/diagram
- choose the appropriate graph/diagram to represent data
- conduct a survey using digital tools
- use a spreadsheet to analyse data and visualise it using a graph

03 Over the course of the lesson

Lesson One

Introduction (5 min): The pupils watch the short, interesting Youtube video adapted to their level:

Public opinion polls

https://www.youtube.com/watch?v=ubR8rEgSZSU

Discussion (10 minutes): The key elements of the video (asking questions, selecting people who represent the population, sample, observing public opinion etc.) are discussed in plenary.

Create your own survey (30 min): Students set up their own survey. They choose a target population (pupils in the class, at school, friends, the whole population of their village, country, etc.). They then formulate (at least) 6 questions aimed at finding out what their target population thinks about a given topic. For each question, they must think about whether it's a

- multiple choice,
- open response,
- opinion
- ...

They also have to think about any possible answers (M1). Students then create a survey in Microsoft Forms (https://forms.office.com/). Any other software is also permitted. We opted for Microsoft Forms because students and teachers have access to it via **Microsoft 365 Copilot** (formerly **Office 365)**. There are three options:

- Students can conduct their own surveys.
- They can use a survey made by us:

Microsoft Forms survey template

Link to Microsoft Forms survey

- They can adapt the survey we did.

Technical explanations of the survey are given in M2.

Homework: Finish creating the survey and get friends, schoolmates, family, etc. to fill it in.

Second lesson

Introduction (5 min): The students are given 5 minutes to contemplate the data they have collected.

Data analysis (10 min): Discuss in plenary the data collected by the students: does all the data correspond to the expectations of the question? Could or should the questions have been asked differently? Should the possible answers have been formulated differently?

Differentiation: for advanced students, a discussion on possible *outliers* can take place, as well as a reflection on what to do with this data.

Correcting the survey (5 min): The students go back to table M1 (M3) and reformulate the questions in their survey.

Research question (15min): Divide the class into teams of 3 or 4 pupils. We recommend favouring teams that work well together and not multi-level: weak together and strong together[3]. This encourages discussion. Each team thinks about an interesting statistical research question that meets the following conditions:

1. impossible to answer yes/no,
2. obligation to combine two survey questions,
3. interesting question for the students.

Here are a few examples based on our survey:

- Who spends more time in front of the screen, girls or boys?
- Do those with an iOS device spend more time in front of the screen, or do others spend more time in front of the screen?
- What is the link between a person's age and their main activity on their smartphone?
- Do people with IOS, Android or Windows smartphones have more TVs at home?

Research question 2 (10 min): Pool (in a google doc) all the possible statistical research questions and choose one of them (discussion + vote). The discussion can focus on (among other things)

- interest,
- relevance,
- formulation (clarity).

Voting may be anonymous or by a show of hands.

Lesson three

Introduction (5 min): Repetition of the research question chosen by vote in the previous lesson.

Result (15 min): Resolution of the issue within the team.

[3] Some recent research in mathematics didactics has shown that in mixed or multi-level groups, stronger students tend to dominate team discussions and make important decisions about the mathematics to be worked on, while weaker students are more involved in more procedural tasks (taking notes, etc.).

Instructions for students :

1. Watch out for *outliers* (see above).
2. Find an answer to the question using the data.
3. Make a graph.

Instructions for teachers :

1. walk around and see what the students are doing
2. signal to a team if they do not consider *outliers*, work with inadequate data, etc.

Formulation of results (15 min): The students formulate an answer to their research question by making a graph and writing a short text. Students can use the graph generated automatically by Microsoft Forms. They can also export the answers to Excel (see M3) and then create a custom graph. Any other program or spreadsheet can also be used.

Presentations I (10 min): Each team presents its results on the board (choice: spokesperson or all together), with their graph in support. They are also asked to present their analysis method.

Warning: It's not enough to give the final result (e.g. "people who own an iOS device also own more TVs at home"), but you need to add an explanation (e.g. "because on average, people who own an iOS device own 2.3 TVs, whereas the others only own 1.8").

The other pupils can react to the results of the other teams and ask questions. The teacher comments on the graph (colours used? Axis labels? Titles? Clear?).

Lesson four

Presentation II (10 min): Finish the presentations started in lesson three.

Wrap up (15 min): During the plenary session, the teacher guides the pupils through a comparative analysis of the different strategies employed by each group. The teacher explicitly emphasises that different answers may all be valid in this context.

The teacher points out that each method has its own strengths and weaknesses. He points out that this module is distinguished by the absence of a single, definitive solution.

If time allows, the teacher can orchestrate a metacognitive discussion to get the students to think about this particularity of statistics: why does it accept several correct answers, unlike the mathematics they usually practise at school? This reflection will enable them to deepen their epistemological understanding of the disciplines and develop their critical thinking when faced with different forms of scientific reasoning.

04 Differentiation possibilities

The teacher will note that this unit allows for natural differentiation adapted to the different levels of students. For the more advanced learners, it is advisable to encourage them to deepen their statistical thinking by exploring more complex concepts such as the identification and treatment of outliers. The teacher can guide them in questioning the impact of these values on their results, the appropriate methods for detecting them and the different strategies for dealing with them (inclusion, exclusion, data transformation).

The intrinsic flexibility of this module is one of its main pedagogical assets. By allowing students to design their own survey and formulate their research questions according to their abilities and interests, the module naturally adapts to the different skill levels present in the classroom.

This open structure offers several advantages:

- Students who need more support can work on more accessible issues.
- More advanced students can engage in more sophisticated statistical analysis.
- Each group can progress at its own pace while participating in the same overall project.
- The teacher can differentiate his or her support according to the specific needs identified.

This automatic differentiation built into the very design of the activity enables the teacher to create an inclusive learning environment where each student is stimulated to his or her optimum level of development.

05 Further criteria to be met as part of the lesson series

a. **Luxembourg context:** The MIDAS team at the University of Luxembourg, led by Prof. Christophe Ley, conducts research into modelling, data science, applied mathematics and statistics. Among other things, they carry out interdisciplinary projects on statistics and sport.

b. **Differentiation:** As described in the previous paragraph, the module contains several levels of differentiation, both in the choice of version of the survey (new survey, adaption or copy) as well as in the choice of the research question chosen.

c. **Reference guide for media literacy[4] :**

   Competence 1: 1.1 Filtering and researching data, information and digital content, 1.2 Analysing and assessing data, information and digital content, 1.4 Processing data, information and digital content.

   Competence 2: 2.1 Working with others, 2.2 Sharing and publishing data, information and digital content

   Competence 3: 3.3 Modelling, structuring and coding

d. **4C model:** communication, collaboration, creativity, critical thinking: The 4Cs are used in this module. The various questions are solved in groups, which requires communication and collaboration on the part of the students. The main idea behind this module on data and statistics is to develop critical thinking skills.

e. **Link to mathematical research:** Data science represents one of the fastest-growing areas where mathematics is being applied to solve modern problems, from predicting disease outbreaks to improving social media algorithms.

[4] https://edumedia.lu/wp-content/uploads/2024/12/Medienkompass_EN_web.pdf

06 Detailed planning of the lesson

Duration	Phases	Focus	Social forms / Methods	Materials and Media	Learning processes
First hour of teaching					
5'	Getting started	What are opinion polls for?	Individual or group work	Tablets or computer	Pupils... ... are familiar with the concept of surveys. ... are familiar with statistical vocabulary.
10'	Discussion	Surveys and statistics	Plenary debate	/	Pupils... ... understand the concept of surveys. .... are familiar with specific vocabulary (population, sample, survey).
30'	Project	Drawing up your own survey	Individual work	M1 & M2 & M3	Pupils... ... are able to think about survey questions. ... are capable of establishing a survey. ...are able to choose a suitable population for their survey. ... are able to use digital tools to draw up a survey.
Second hour of teaching					
5'	Getting started	Data	Individual work	Tablets or computers	Pupils... ... are familiar with the concept of data. ... are able to consider data in a computer program.
10'	Analysis	Data	Plenary debate	Tablets or computers	Pupils... ... are able to interpret data. ... are able to identify outliers.
5'	Reflection	Survey	Individual work	M1 & M3	Pupils... ... reflect on their own survey and correct it.
15'	Research I	Research question I	Group work	Tablets, computers or paper	Pupils... ... are able to establish a research question.

10'	Research II	Choosing a research question	Plenary debate	/	Pupils... ... are able to choose a relevant research question.
Third hour of teaching					
5'	Getting started	Research question	Plenary debate	/	Pupils... ... know the research question to be studied.
15'	Resolution	Data interpretation	Group work	Tablets or computers	Pupils.... ... are able to combine data and interpret it. ... are able to produce a graphical representation.
15'	Formulation	Formulation of the result	Group work	Tablets, computers or paper	Pupils... ... are able to formulate their conclusions in writing.
10'	Presentation	Presentation of results	Group work	Projector	Pupils... ... are able to present their findings to the class using a graph.
Fourth hour of teaching					
10'	Presentation	Presentation of results	Group work	Projector	Pupils... ... are able to present their findings to the class using a graph.
15'	Wrap up	Different strategies	Plenary debate	/	Pupils... ... know that there are several different strategies for solving a research question in statistics. ... understand that there is no single answer in statistics.

4.3Teaching materials

M1 Creating a survey

Survey theme		
Target population		
Questions	Formulation	Format
1		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
2		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
3		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
4		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
5		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
6		☐ Multiple choices ☐ Open answer ☐ Opinion ☐ other

M2 Microsoft Forms

Microsoft Forms is a powerful tool that makes it easy to create surveys, quizzes and questionnaires. This application is available to all teachers and pupils with a **Microsoft 365 Copilot** account (formerly **Office 365**).

Method 1: Create a survey from scratch

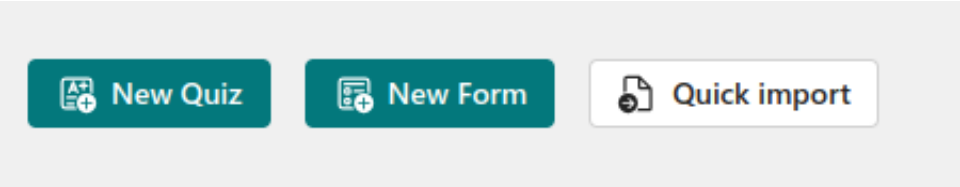
To set up a survey from scratch in Microsoft Forms, follow the steps below.

Step 1: Log in

Go to https://forms.office.com/ and log in with your Microsost 365 Copilot (formerly Office 365) credentials.

Step 2: Create a new form

Click on the "New Form " button to begin.

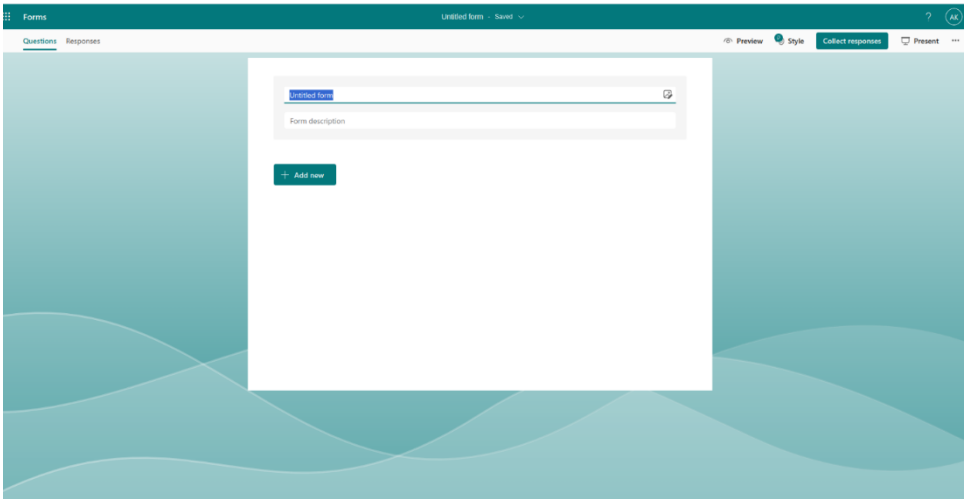


Step 3: Customise your form

- Give your survey a clear title.
- Add a detailed description to explain the purpose of the survey.
- You can also customise the theme and appearance.

Step 4: Add questions

Click on "Add New" to formulate a question.



Step 5: Choose the type of question

Select the format best suited to your question:

- Multiple choice
- Text
- Rating

- Date
- Likert scale
- And many more...

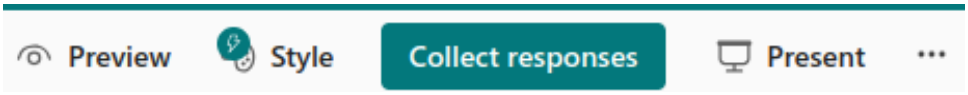
Step 6: Preview your survey

Once you have created your questions, use the "preview" option to see what the survey looks like.

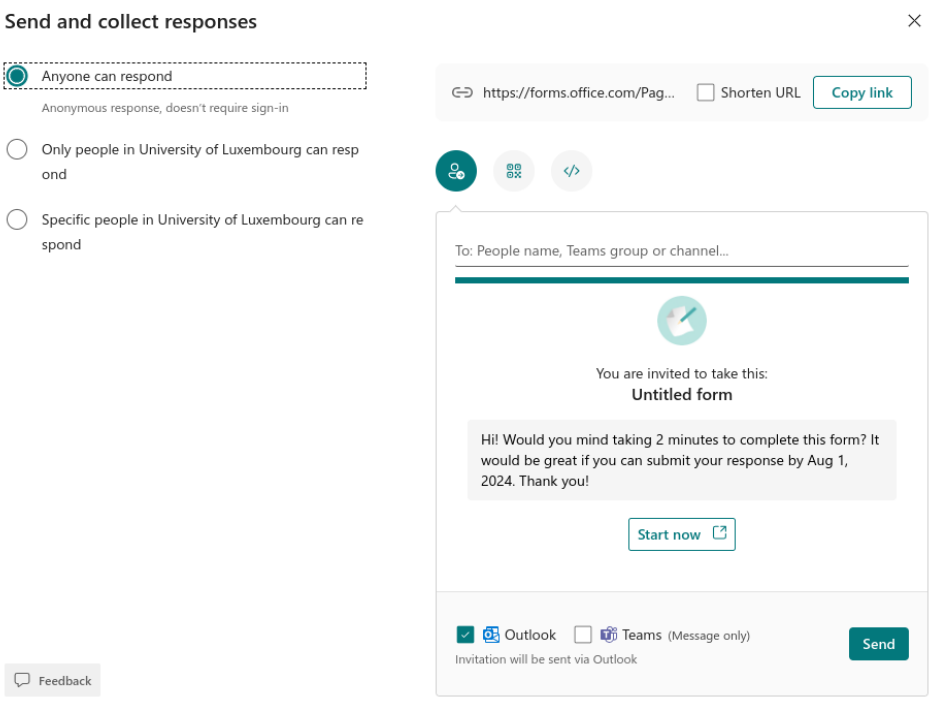
Step 7: Configure sharing parameters

To maximise participation :

- Go to "Collect responses".



- Select "Anyone can respond".



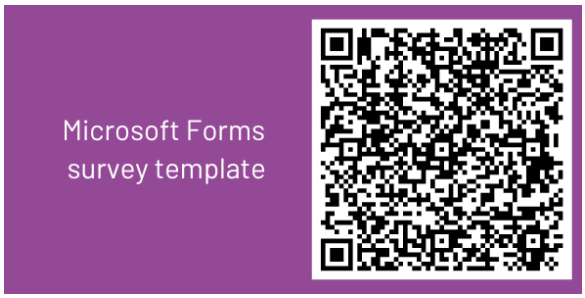
Step 8: Distribute your survey

Copy the link or QR code to share your survey with participants.

Method 2: Use the pre-established survey.

Step 1: Access the model

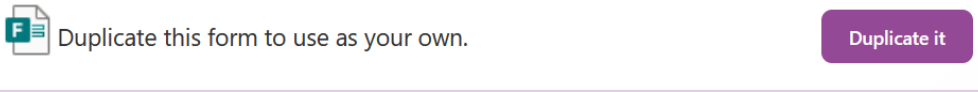
Open the template link :



Link to Microsoft Forms survey

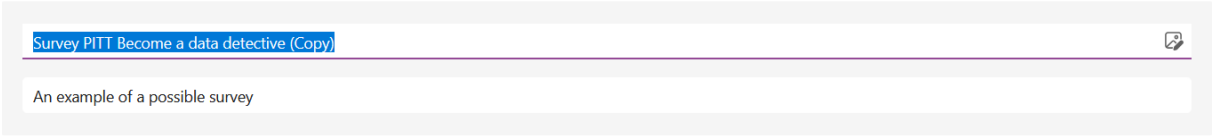
Step 2: Duplicate the model

Click on "Duplicate" to create your own customised copy.



Step 3: Customise the form

- Modify the title and description to suit your needs.
- Customise the appearance to suit your preferences.



Step 4: Adapt the questions

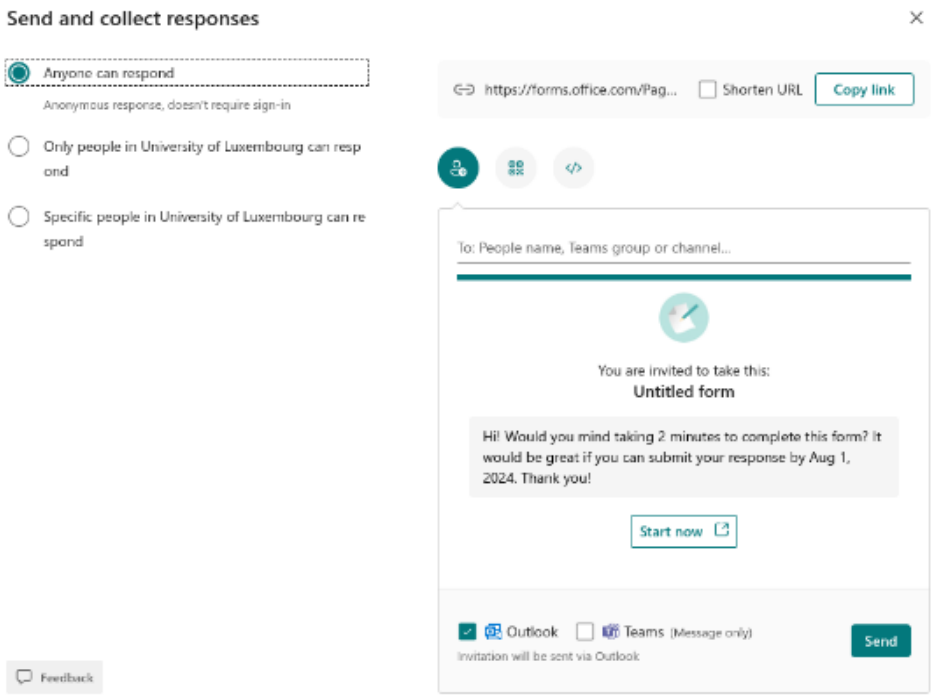
Modify existing questions or add new ones according to your specific needs.

Step 5: Configure sharing parameters

As with method 1:

- Go to "Collect responses".

- Select "Anyone can respond".



Step 6: Distribute your survey

Copy the link or QR code to share your survey with participants.

M3 Rethinking your survey

Survey theme		
Target population		
Questions	Formulation	Format
1		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
2		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
3		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
4		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
5		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other
6		☐ Multiple choice ☐ Open answer ☐ Opinion ☐ other

4.4 Interdisciplinary ideas

01 Life and Society (VIESO)

This statistics module provides an excellent opportunity to develop pupils' analytical skills in an interdisciplinary context. By discovering that several statistical interpretations can be valid, pupils are naturally led to think about the wider implications of this interpretative flexibility.

The teacher can exploit this opening to establish significant connections with the Life and Society course, particularly its Culture and Communication strand. The fundamental questions addressed in this strand find a direct echo in statistical analysis:

- What is the difference between prejudice, judgement and knowledge?
- What are the media and how are they used?
- Are the media neutral? Do the media show us the world or just images of the world?
- Information or propaganda?
- Am I letting myself be manipulated by the media?

To enrich this interdisciplinary connection, teachers can use the following video from the French educational platform Lumni[5] . This clip will enable pupils to make the link between their statistical work in class and the wider issues of media literacy and critical thinking in the information society.

https://www.lumni.fr/video/les-journalistes-sondeurs

This integrated approach not only strengthens pupils' mathematical skills but also develops their ability to make informed judgements about quantitative information presented in the media and in everyday life.

Remark: While the video is presented in French, you'll find the full transcript beneath it, which you can easily translate to English using online translation tools.

[5] Lumni is the educational offering for pupils and teachers from all the players in the public audiovisual sector (Arte, France Médias Monde, France Télévisions, INA, Radio France and TV5Monde). This unique, free, expert and commercial-free offering provides access to culture, knowledge and learning for all children aged 3 to 20 and covers all school subjects from nursery school to the end of secondary school.

02 PITT module: Data Viz Superpowers

The integration of the PITT Data Viz Superpowers module offers a natural and enriching extension to the initial statistical work. This approach would deepen the pupils' learning experience by taking them from analysis to visual communication of data.

There are a number of ways in which teachers can integrate this:

Transforming surveys into visual storytelling - The data collected by pupils in their initial survey could be reworked into storytelling (Exercise 11 of *3#Data Viz Superpowers 3.3 Learning Materials*). This approach would enable pupils to understand how statistics can be integrated into a coherent and persuasive narrative.

Creative data representations - Beyond conventional graphs, the teacher could encourage pupils to explore more creative and innovative visualisations (as suggested in *3#Data Viz Superpowers 3.5 Assessment Ideas*). This would stimulate their creativity while enhancing their understanding of the data.

Critical analysis of misleading visualisations - Exercises 12 to 15 *in 3#Data Viz Superpowers 3.3 Teaching Materials* provide a valuable opportunity to examine graphic manipulation techniques. Teachers could use these resources to develop pupils' critical thinking skills in relation to potentially misleading visual representations in the media.

This integrated approach would not only enhance technical skills in data visualisation, but also provide an opportunity to reflect more deeply on the ethical and communicational dimensions of statistics. Pupils would discover the dual nature of visualisations: powerful tools for communication but also potential instruments for manipulation, developing their visual literacy and their ability to navigate in a world where data is omnipresent.

4.5 More on this topic

01 A short history of statistics

The history of statistics began in the 17th century with William Petty's *Down Survey*. This cadastral survey of Ireland, carried out in 1655-1656, was intended to facilitate the precise redistribution of Irish land confiscated by the English. The name *Down Survey* is thought to have been chosen by Petty either because the results were plotted on maps, or in reference to the surveyor's chain that had to be "down" during the measurements. At the time, this work represented one of the most accurate cartographies and was the first British imperial topographical survey of an entire conquered territory. (Wikipedia)

This founding example perfectly illustrates the birth of statistics: a tool developed to meet the State's need to collect data in order to administer the population and the economy. This dimension is reflected in the very etymology of the word "statistics", derived from the Latin "statisticum" (affairs of state), a term introduced into German by Gottfried Achenwall in the form "Statistik". Over the course of the 19th century, these methods diversified and were applied to many fields, considerably broadening their scope of application. (Vogt & Osipenko)

Statistical understanding developed gradually. Probability theory was first explored by Christiaan Huygens in 1657, followed by Roger Cotes' error theory in 1722 (published posthumously). Major advances were made by a number of scientists, including Jakob Bernoulli, Abraham de Moivre, Thomas Simpson and Pierre-Simon Laplace, culminating in Andrei Kolmogorov's probability axioms in the 20th century.

02 The lady with the lamp

Florence Nightingale, famously known as the "Lady with the Lamp", was much more than just a nurse. A talented mathematician and statistician, she was also a pioneer in the fields of data science and data visualisation.

Born into a wealthy family on 12 May 1820, Florence received a careful education under the supervision of her father. He taught her not only subjects appropriate for Victorian women, such as drawing and needlework, but also traditional academic subjects. From an early age, Florence showed an exceptional passion for learning.

Her early interest in statistics can be seen in a letter to her grandmother, in which the young Florence meticulously listed the animals she had observed during her visits to the Zoological Society, already demonstrating her taste for collecting and organising data:

Dear Grandmama. The baby is pretty. I have been to the Zoological Society twice. There are 2 leopards, 2 bears, 2 parrots, 2 emeus (which are very large birds), 2 rabbits. 1 lion, 2 cockaatoos, 3 squirrells, 4 kanguroos, 6 monkies, (3 in a cage, 3 chained to a pole with a little house at the top). 1 rattel, (a very fierce creature), several Esquimaux dogs, Captain's Parry's Esquimaux dog, 1 guinea

pig, 1 Costi Monti, 3 llamas, (1brown one, 1 white one, & a small brown one), & other creatures that I forget the name of.[6]

Although she excelled in all subjects, her real passion remained mathematics. Her sister Parthenope testified to this devotion when she wrote: (Bostridge, 2008):

Florence has taken up mathematics - and as with everything she does, she puts her heart and soul into it and works hard.

Rejecting the conventional fate of a housewife, Nightingale aspired to a profession related to statistics. To circumvent the restrictions imposed on women of her time, she chose to study nursing in Germany, one of the few acceptable professional paths for women. When the Crimean War broke out in 1853, she seized the opportunity to put her skills into practice and left as a nursing sister a year later.

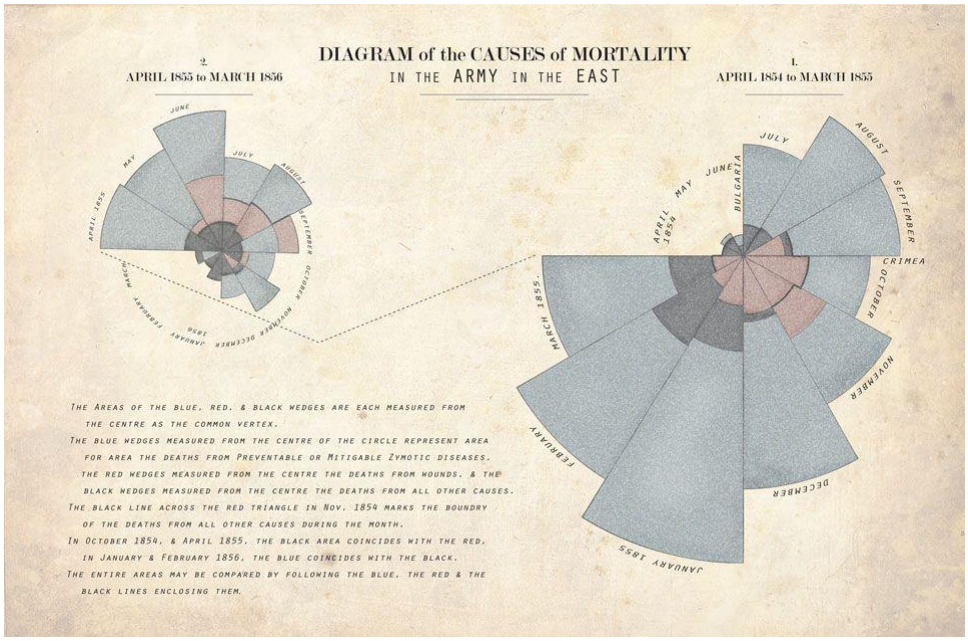
During her two years in the Crimea, Nightingale undertook a radical transformation of military hospitals while supervising the work of the Royal Commission on Army Health. Working closely with several statisticians, she wrote numerous reports and articles. It was during this period that she earned her famous nickname, when The Times reported that she walked the wards at night, lamp in hand, looking after the wounded.

Source: https://www.berliner-zeitung.de/gesundheit-oekologie/die-lady-mit-der-lampe-florence-nightingale-corona-pandemie-krankenpflege-li.83282

His main mission was to highlight the deplorable conditions of soldiers and the shortcomings in the organisation of military statistics. Nightingale rigorously collected the relevant data, interpreted it accurately and developed innovative visualisations to communicate her findings effectively.

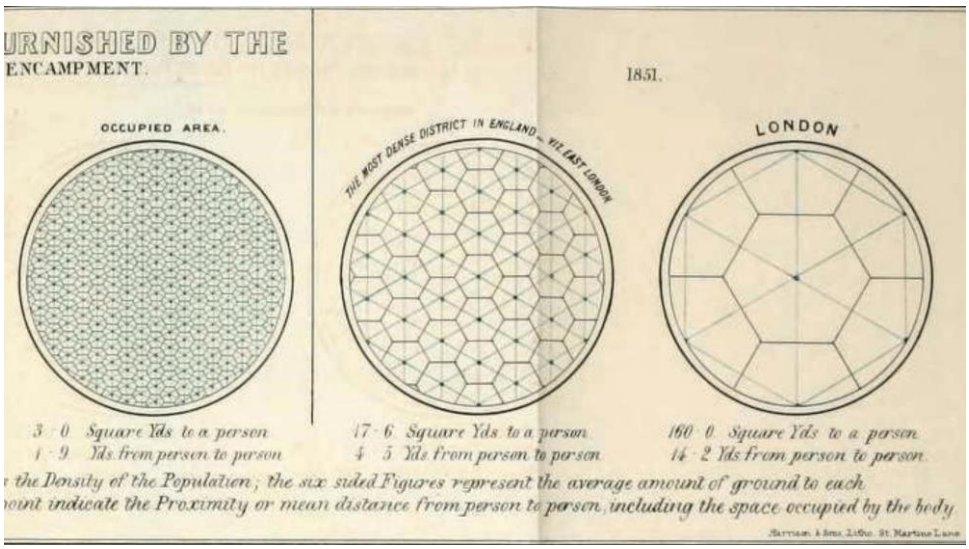
Diagrams are extremely useful for understanding certain vital statistical issues.

[6] *Nightingale, Florence. 1828. Letter to Grandmother.* https://cwfn.uoguelph.ca/



Source: (Bradshaw, 2017)

Her diagrams revealed the appalling mortality in the British Army, identifying overcrowding, poor ventilation and unsanitary barracks as the main causes. To demonstrate the extent to which barracks were overcrowded, Nightingale created the following visualisation.



Source: (Bradshaw, 2017)

Faced with the chaotic medical statistics of the British army, which she described as being "in great confusion", Nightingale pleaded tirelessly for:

- Correct use of data
- The need to report information accurately
- Uniform collection of hospital statistics

Thanks to her perseverance, the army's medical statistics were finally reorganised and a medical corps statistics channel was established.

Florence Nightingale remains a visionary figure whose impact on hospital reform and healthcare systems was considerable. Her pioneering approach, using diagrams to tell stories based on data, was revolutionary in its day, particularly for a woman. With her mastery of visualisations and rigorous data analysis, she can rightly be considered one of the first women to practise what we now call data science, laying the foundations for what is now an essential discipline.

For those interested, we refer you to Sarah Hart's interesting lecture, *The Mathematical Life of Florence Nightingale*:

Florence Nightingale

https://www.gresham.ac.uk/watch-now/maths-nightingale

03 Known paradoxes

Navigating the world of statistics can be complex, particularly when figures are presented in the media, politics, the economy or health, sometimes creating paradoxical situations. It is essential to examine these data carefully to avoid any misinterpretation. In the following video the famous Simpson paradox is explained.

How statistics can be misleading

https://www.youtube.com/watch?v=sxYrzzy3cq8&t=41s

The following, more concise video, produced by Lucie Zeches and Jean-Paul Bertemes for the *Ziel mer keng* series (with English subtitles), sheds light on the common statistical errors we encounter on a daily basis.

Link Ziel mer keng

04 Statistics and sport

The MIDAS team (Modelling, Interdisciplinary research, Data science, Applied mathematics and Statistics) at the University of Luxembourg, led by Professor Christophe Ley, aims to develop innovative statistical and machine learning procedures based on new mathematical and computer tools to meet the challenges posed by today's increasingly complex and voluminous data sets.

According to legend, King Midas died because everything he touched turned to gold, so he couldn't eat or drink. With data, the same thing can happen: having tons of data may seem great at first glance (like King Midas' ability), but if you don't know what to do with it, it will mislead you or you simply won't be able to do anything with it. The MIDAS team wants to provide and explain methods that prevent you from ending up with data like King Midas.[7]

One of the MIDAS team's key themes is sports analytics. Sports analytics is the collection of relevant historical statistics that can provide a competitive advantage to a team or individual by helping to inform players, coaches and other staff and facilitate decision making during and prior to sporting events.[8]

Prof. Christophe Ley has published a book on the subject of sports statistics, which deals with the most diverse and fruitful combinations of data analysis and sport, such as a study conducted by the LIH on running shoes that reduce injuries while running, or image recognition techniques for more effective match analysis.

In 2018, Christophe Ley, a professor of statistics at Ghent University (Belgium) at the time, combined statistics and sport for the first time. Together with colleagues, he developed a new method for predicting the winner of the 2018 World Cup.

Unlike traditional methods used by bookmakers that rely primarily on human analysis of betting and performance data, Ley's approach combines statistics with machine learning to create a more reliable model.

The core of their method is based on the use of the "Random Forest", a machine-learning algorithm consisting of decision trees that are constantly run through by a computer programme. The decision parameters relating to players, opponents and the environment are defined partly

[7] Description taken from the MIDAS web page.
[8] https://en.wikipedia.org/wiki/Sports_analytics

by machines and partly by human beings, and the World Cup is replayed again and again. As each replay is associated with an (automatic) learning process, predictions become increasingly reliable (Hauser, 2018).

Recently, Christophe Ley, accompanied by members of his MIDAS team, visited the facilities of Borussia Dortmund, a German football club, just before their Champions League match against Paris. The visit was part of the *Math4Sports* project, which brings together several major European football clubs: Arsenal FC, Athletic Bilbao, PSV Eindhoven, Benfica Lisbon and Borussia Dortmund.

The main aim of this collaboration is to use statistics and data analysis to improve injury prevention in football. At a meeting in Dortmund, university researchers and sports scientists discussed how science could help reduce the risk of injury, a recurring problem for BVB in recent seasons.

Statistics are increasingly used in modern sport: players' movements are tracked during training and matches, generating huge amounts of data that statisticians can analyse to optimise performance and prevent injury.

A concrete example of the usefulness of sports statistics is another collaboration by the MIDAS team with the Doneck Dolphins Trier, a wheelchair basketball team. Ley and his colleagues have developed a method for predicting player performance and proposing optimal team compositions, taking into account the specific rules concerning handicap points in this sport (Bertemes, 2024).

05 Statistics and monitoring for tomorrow's schools

LUCET (Luxembourg Centre for Educational Testing) at the University of Luxembourg is an educational research group whose main mission is to implement, improve and ensure school monitoring in Luxembourg, an activity that mobilises the majority of its resources.

Founded jointly by the University of Luxembourg and the Ministry of Education in July 2014, this unique research and transfer centre is the culmination of the "Standardised Tests" (ÉpStan) project. This long-term research programme (2007-2014), funded by the Luxembourg Ministry of Education, was already devoted to school monitoring and analysis of the country's educational challenges.

Given the importance of the issues identified, the 2014-2017 Pluriannual Establishment Contract between the State and the University officially integrated this programme into the university structure, creating a research centre attached to the Faculty.

Standardised tests (EpStan) are currently used as school monitoring tools in Luxembourg and consist of tests and questionnaires presented in pencil and paper versions, and/or on computer. These tests provide a standardised assessment of skills in key areas of school education or key aspects of the quality of teaching, the school and classroom climate and pupils' motivation to learn. More specifically, once a year, at the start of the new compulsory schooling learning cycle, i.e. cycles 2.1, 3.1, 4.1 and 7^{e}and 5^{e}of the ESC/ESG, EpStan enables all pupils at the corresponding levels of study to be assessed as to whether the learning objectives of the previous learning cycle have been achieved. Every year, all the pupils in each class concerned take part in EpStan. This represents around 28,000 pupils per year. To ensure a fair comparison of performance, EpStan

systematically takes account of pupils' socio-economic and socio-cultural background, and thus of characteristics that have been shown to have a major influence on school success on average.

Luxembourg faces a major educational challenge: managing an increasingly diverse pupil population effectively. According to the OECD's PISA studies, this challenge is not unique to the country but is particularly pronounced due to specific factors such as its small size, its central position in Europe, its multilingualism and its economy based on immigration.

This situation makes Luxembourg a veritable "living laboratory", where demographic changes occur more rapidly than elsewhere. The solutions developed in Luxembourg to offer equal opportunities for success to all pupils, regardless of their socio-economic, cultural or linguistic background, could therefore serve as a model for other countries facing similar challenges in the decades to come.

It is in this context that Standardised Tests (EpStan) play a crucial role. These evaluations provide relevant data that make a major contribution to educational management and empirical research in the field of training. They make it possible to objectively evaluate the performance, equity and evolution of the Luxembourg school system over the long term, thus providing a solid scientific basis for responding to the challenges of diversity.

While assessments for all elementary schools are paper-based, EpStan assessments for secondary schools are fully computerised and web-based (using OASYS, LUCET's in-house online assessment system).

Thanks to Luxembourg's school tracking programme, LUCET not only provides timely and relevant information to national education stakeholders but also constitutes a unique and incredibly rich longitudinal database on the evolution of pupils' skills profiles, their school career and, eventually, their life course.

References

1. Andrews, R. (2021). The radiant diagrams of Florence Nightingale Michael Friendly.
2. Bertemes, J-P. (2024). BVB: Wird ein luxemburgischer Forscher Borussia Dortmund in Zukunft helfen? Science.lu. https://science.lu/fr/node/16698
3. Bostridge, M. (2008). *Florence Nightingale*. London: Penguin Books.
4. Bradshaw, NA. (2017). Florence Nightingale (1820-1910): A Pioneer of Data Visualisation. In: Beery, J., Greenwald, S., Jensen-Vallin, J., Mast, M. (eds) *Women in Mathematics*. Association for Women in Mathematics Series, vol 10. Springer, Cham. https://doi.org/10.1007/978-3-319-66694-5_11
5. Hauser, S. (2018). Luxemburger Forscher weiß (ziemlich sicher), wer Fußball-Weltmeister wird. *Science.lu*. https://science.lu/fr/treffsichere-prognose/luxemburger-forscher-weiss-ziemlich-sicher-wer-fussball-weltmeister-wird
6. Luxembourg Centre for Educational Testing (LUCET), https://www.uni.lu/fhse-en/research-groups/luxembourg-centre-for-educational-testing-lucet/
7. Vogt A., Osipenko M. What is statistics? - From a historical perspective (SE).

4.6 A word from the scientists

Christophe Ley is Associate Professor of Mathematics at the University of Luxembourg, where he heads the *Modelling, Interdisciplinary Research, Data Science, Applied Mathematics and Statistics* group within the Mathematics Department.

A specialist in applied mathematics, his research focuses on statistics and data science, with a strong emphasis on interdisciplinary approaches.

He holds a number of important positions, including the presidency of the Statistical Society of Luxembourg and of ECAS (European organisation for Advanced Courses in Statistics). He is also head of the international S-TRAINING network, which develops new data science methods for sports analysis and sports medicine.

Christophe obtained his doctorate from the Université libre de Bruxelles in 2010. He did most of his studies at the same university, with the exception of his first year at the University of Luxembourg when it was set up in 2003-2004. After completing his doctorate, he continued his post-doctoral research at the Université Libre de Bruxelles, before taking up a position as associate professor at Ghent University. He then returned to his native Luxembourg to continue his academic career.

To find out more about Christophe Ley and his research, don't miss the excellent SciLux season 3 podcast with Hanna Siemaszko, which offers an in-depth look at his background and work.

https://scilux.buzzsprout.com/1412332/episodes/11997749-season-3-episode-8-statistics-and-data-science

ISSN 2738-9383