Software for learning and for doing statistics and probability – Looking back and looking forward from a personal perspective

Software para aprender y para hacer estadística y probabilidad – Mirando atrás y adelante desde una perspectiva personal

Rolf Biehler

Universität Paderborn, Germany

Abstract

The paper discusses requirements for software that supports both the learning and the doing of statistics. It looks back into the 1990s and looks forward to new challenges for such tools stemming from an updated conception of statistical literacy and challenges from big data, the exploding use of data in society and the emergence of data science. A focus is on Fathom, TinkerPlots and Codap, which are looked at from the perspective of requirements for tools for statistics education. Experiences and success conditions for using these tools in various educational contexts are reported, namely in primary and secondary education and pre-service and in-service teacher education. New challenges from data science require new tools for education with new features. The paper finishes with some ideas and experience from a recent project on data science education at the upper secondary level

Keywords: software for learning and doing statistics, key attributes of software, data science education, statistical literacy

Resumen

El artículo discute los requisitos que el software debe cumplir para que pueda apoyar tanto el aprendizaje como la práctica de la estadística. Mira hacia la década de 1990 y hacia el futuro, para identificar los nuevos desafíos que para estas herramientas surgen de una concepción actualizada de la alfabetización estadística y los desafíos que plantean el uso de *big data*, la explosión del uso masivo de datos en la sociedad y la emergencia de la ciencia de los datos. Se centra en Fathom, TinkerPlots y Codap, que se analizan desde la perspectiva de los requisitos que deben cumplir las herramientas para la educación estadística. Se informa de experiencias realizadas con éxito y condiciones de uso de estas herramientas en varios contextos educativos, tanto en educación primaria como secundaria y la formación inicial y en servicio de los profesores. Los nuevos desafíos que plantea la ciencia de los datos requieren nuevas herramientas para la educación con nuevas características. El artículo termina con algunas ideas y experiencias de un proyecto reciente sobre educación en ciencia de los datos en el nivel de bachillerato.

Palabras clave: software para el aprendizaje y la práctica estadística, atributos clave del software, educación en la ciencia de los datos, alfabetización estadística

1. Introduction: A brief look backwards

Discussing software under the perspective of how it supports both the learning and the doing of statistics has recently been resumed by Amelia McNamara (2015). She kindly refers to my paper Biehler (1997) where this topic was discussed under certain historical circumstances. The current paper starts with looking back into the origins of the 1997 paper, where a vision of future software tools for statistics education was elaborated. The programs Fathom, TinkerPlots and Codap fulfil many of the envisioned features but have to be also discussed from the perspective of using them in various educational contexts.

The 1997 paper was based on a talk given at ICoTS 4 in Marrakech 1994 and was further inspired by the conference that Carmen Batanero organised in Granada 1996 on technology in statistics education (Garfield & Burrill, 1997). It was based on several

Biehler, R. (2019). Software for learning and for doing statistics and probability – Looking back and looking forward from a personal perspective. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en www.ugr.es/local/fqm126/civeest.html

assumptions and visions about what should be the content and working style in probability and statistics education. A little clearer title would have been: One software for both goals: for learning and for doing statistics. Contemporary changes in the field of statistics such as the emergence of Exploratory Data Analysis (EDA) (Tukey, 1977) and the growth of computational inferential statistics with its heavy use of simulations (Efron, 1982) were reflected in this paper. EDA puts the interactive, heavily graphical exploration in its focus and works with multivariate data sets. Probability modelling could be extended to new fields as computational tools allowed to overcome the use of models with only the relatively simple assumptions that were trackable in an analytical way. More complex and realistic models can be studied through computational modelling. Models can be related to real data so that models can be validated and updated in the next step of a modelling cycle.

If we want to implement these new practices in introductory education, we need adequate computational tools similar to those used in practice. On the other hand, computational technology has the potential to support active learning by interactive visualisations, by providing environments for experimentation with methods and by use of simulations that make probability a lively experience for students. Several programs were around – that we would call applets today – and the question was: Can we imagine a tool for doing statistics that at the same time can support the creation of interactive applets that support active learning. Based on this approach the question was, which features should a computational tool have so that it can support doing these new practices in introductory statistics and probability education in a model-like way in courses at high school and at the entrance level of universities. Within the context of statistics itself, different kinds of tools had been developed aiming at supporting the new statistical practices since the 1980s. On the one hand, the programming language S had already been developed (Becker, Chambers, & Wilks, 1988), on the other hand, Data Desk was the new prototypical tool for doing exploratory statistics with a graphical user interface (Velleman, 1989). In the beginning, DataDesk was only available on Apple Macintosh computers. The Windows operating system was developed not before the 1990s. S has now been replaced by R, which has become the standard tool for doing statistics and doing research on statistical methods. From the currently available tools, JMP is similar to Data Desk in interaction style, while Data Desk is still available. Both tools, JMP and R, are of course much more advanced concerning the statistical and graphical methods they incorporate. Moreover, the user interface has been elaborated as well. Both these recent tools have educational uses, see, e.g. Kraft (2016) for JMP and Gould et al. (2016) for R. The educational uses were partly done with adaptations of these tools to fit students' needs and skills better.

2. Background of the 1997 paper about a vision of software for doing and for learning probability and statistics

The 1997 paper resulted from a failure of the curriculum and software development project MEDASS (Modelling in connection with EDA and stochastic simulation). The project's objective was the development of curriculum material for mathematics, history, political science and geography in secondary education, which was to have cross-references to each other. The mathematics teacher should find context information for the data used in the mathematics classroom and the geography, social and political science teacher can get background information on statistical and graphical methods that she uses in her classroom by referring to the material for the mathematics teacher. These materials were published (Kohorst, 1992; Noll & Schmidt, 1994; Portscheller, 1992) In parallel, requirements for a software tool that would support cross-curricular exploratory data analysis was developed by the project team. The software company that was hired to develop the software initially agreed to realise the specification but, in the end, withdraw from the agreement. We took this as an encouragement to publish our requirements first as a 200 pages extended specification of the MEDASS software conception (Biehler & Rach, 1992), and the 1997 paper was a summary of significant points, hoping that this paper may influence future developments, which it seemingly did. As a master thesis in computer science a much simpler version of our specifications, which we called MEDASS light, was developed in the late 1990s (Bauer, Biehler, & Rach, 1999).

The available tools that shaped our thinking were - on the professional side - with a command language interface the tools S, Splus (later R) and ISP-PC, developed by Peter Huber for PCs in the late 1980s (Huber, 2000). We also looked at DataDesk and Statview with a Graphical User Interface (only available for Apple Mac computers in those days). A further tool that we considered at was Survo 84c (Mustonen, 1992; Puranen, 1994), where command language code could be embedded in a text file – thereby creating a so-called “sucro” and selectively executed, a very early version of what today is possible in Jupyter notebooks (Toomey, 2017), which we will discuss below in more depth. On the educational side, we had programs of route-type character and landscape-type character. This distinction was coined by Arthur Bakker (2002) later. The route-type programs include interactive simulations, experiments and visualisations for certain learning goals. Examples include showing the sensitivity of the mean against outlying values or visualising the sum of the squared residuals in a regression context. Among the landscape type of software, we looked at, was DataScope (1994) and the simulation software ProbSim developed by Konold (1994), which were created starting in the late 1980s. Another relevant tool was Tabletop developed by Chris Hancock (1995). However, there was agreement that, in principle, chance (probability) and data should not be separately treated by two different tools, so we imagined a tool that supports both: simulation and data analysis. These requirements were realised in slightly different ways in Fathom and TinkerPlots several years later.

From this analysis, our first basic requirement was a medium size “model tool” that has similarities to professional tools but has a more appropriate learning curve, that covers basic statistical methods and supports an exploratory working style in statistics (in EDA, in inferential statistics and concerning statistical graphs). It should be possible to study the properties of statistical methods by simulation instead of using analytical probability models. A reference was made to Thisted (1986) who distinguished three types of functional environments “Data analysis environment”, “Monte Carlo workbench” and “Theoretical statistician‘s environment”. This model tool would allow to practice statistics and probability simulation similar to the practice outside a school and to learn this practice where tools are indispensable. In this first sense, the model tool also is a tool for learning (the practice of) statistics. The recommendations included to develop a tool with a graphical user interface with direct manipulation of data, have command language input facilities only when needed (such as Fathom later developed the formula editor). We will discuss in the next paragraph how these requirements were realised in tools such as Fathom and TinkerPlots. A fundamental limitation of GUI tools consists in that the actions are not recorded as a sequence of commands, so our vision included a tool having a journal player/recorder similar to the history function of S. However, such a recording or history function was never realised in these tools. A more recent development is iNZight (www.stat.auckland.ac.nz/~wild/iNZight/), which is based on R. Here it would be possible to include such a recording feature and seemingly there are plans to implement such a feature in iNZight while providing simple access for students.

A second basic requirement was that the envisioned tool is also a meta-tool for creating particular purpose learning environments or microworlds. The programming should be done mainly by programming by example such as in spreadsheets, without the usual use of coding. The resulting microworlds can be called "embedded microworlds". Their essential feature is that they can be changed and adapted without specialised knowledge in a specific programming language but with knowledge of the hosting software environment. This adaptability is an advantage over stand-alone microworlds.

3. Tools for learning and doing statistics: visions and reality since the year 2000

Tools such as Fathom (since 2002, fathom.concord.org) and TinkerPlots (since 2005, www.tinkerplots.com) realised many aspects of the 1997 vision and some more, focussing on school education and beginning education at college level. Konold (2007) provides background knowledge on designing TinkerPlots from earlier software origins and software conceptions. My working group has created German localisations of Fathom (2006) and TinkerPlots (2017) (www.stochastik-interaktiv.de). A tool that is based on experiences and features of Fathom and TinkerPlots is Codap (codap.concord.org). It has not yet the functionality of Fathom and TinkerPlots but goes beyond these tools, for instance, in that it provides a modern hierarchical data structure and includes maps and facilities for location-based data. Expert programmers could continuously add Apps and features. Recently, a sampler for elementary simulations in the style of TinkerPlots came under development, and a tool for implementing decision trees has already been added. An attractive feature is that Codap is web-based and provided for free. Other language versions are straightforward to create. In our localisation, we used the terminology we developed for the German localisation of Fathom and TinkerPlots to create a German localisation of Codap.

The development of Fathom, TinkerPlots and Codap would not have been possible without millions of public US money that an excellent team of developers and statistics educators got from the National Science Foundation NSF and other sources. These tools have not been a tremendous commercial success but have been used in many research-based pilot projects in several countries including England, Israel, USA, New Zealand and Australia and Germany. The paper by Biehler, Ben-Zvi, Bakker, and Makar (2013) provides a recent review of existing software tools, learning environments accompanying these tools and related research studies and it updates requirements from software tools for learning and doing statistics. There have been much more developments besides these three tools already mentioned. A tool, which is inspired by the same basic philosophy is iNZight, mentioned above already, which is a free tool. It is based on the programming environment R but provides a graphical user interface similar to Fathom, TinkerPlots and Codap.

4. Tools for learning and doing statistics and probability – their use in educational contexts

The 1997 paper envisioned one tool for many purposes. However, in our research and development projects at the universities in Kassel and Paderborn, we used different tools in different contexts reaching from university-based teacher education, school contexts, and professional development courses for in-service teachers.

University-based pre-service teacher education

Student teachers in Paderborn have to acquire content knowledge and pedagogical content knowledge in probability and statistics. Secondary teachers (grade 5 to 10) have two separate courses on these topics whereas primary teachers (grade 1 to 4) have to attend an integrated course. We are using Fathom in the courses for secondary teachers and TinkerPlots in the courses for primary teachers. These contexts allow semester-long intensive learning and the use of one tool, through which our students can reach moderate expertise in probability modelling and statistical data analysis (exploratory and inferential). Without Fathom (secondary) and TinkerPlots (primary), this achievement is difficult to imagine. However, it turned out that although both tools are comparably easy to learn for students, students need a lot of support and careful integration of learning to use the tool in parallel to learning concepts and methods. For example, we developed the multimedia tool eFathom (Biehler & Hofmann, 2011; Hofmann, 2012), which provides an introduction into the use of Fathom for data analysis and simulation. eFathom is suitable for students' self-regulated learning through four modules, that use texts and video tutorials and related activities with Fathom. Our experience resonates with experiences and studies on the use of digital tools in other domains of mathematics education, where the role of organising a careful instrumental genesis that turns software into a thinking tool for learners through an instrumental orchestration has been widely recognised (Trouche, 2004).

Therefore, success conditions include: the instructors have done a careful analysis of the tool's constraints and affordances, teaching material was tested in design-based research cycles, a careful "didactical orchestration / instrumental genesis" of integrating content and tool learning is implemented, and student difficulties and reasoning with tools have been the object of detailed empirical studies. Moreover, students need support for learning complex multistep activities. For guiding simulations, we developed a "simulation and modelling scheme"(Maxara & Biehler, 2007). For guiding data analysis, we developed ideas about how to keep a "data diary" or how to write brief reports (Biehler, 2005, 2007). The process model of Wild and Pfannkuch (1999) was used in a simplified version. We also implemented ideas how to relate content knowledge and pedagogical content knowledge in a course (Batanero, Biehler, Engel, Maxara, & Vogel, 2005).

Statistics and probability education at school level in Germany

The implementation at school level faces much more difficulties. Many mathematics teachers at school level in Germany do not feel as statisticians or probability modellers and do not intend to teach these activities at great length and depth. Most teachers have not been longing for tools such as Fathom or TinkerPlots that would enable them to do this kind of teaching. Moreover, many think they cannot afford the time for teaching with a tool just suitable for probability and statistics and prefer (if at all) more general tools such as GeoGebra, EXCEL, or graphic calculators from Casio or Texas Instruments (TI Nspire).

National standards in Germany (KMK, 2012, 2004) highlight data and chance as one of the leading ideas, but the curriculum at primary and lower secondary level does not assign much obligatory time for probability and statistics. It is consistent with the curricula, to avoid any extensive work with real data sets and to avoid the use of simulations. The total time devoted to probability and statistics in grade 5 to 10 may be less than one week per year on average. Even if statistical tools are taught, they may not intensely be used for data exploration. For instance, although box plots have become part of most secondary curricula, students most often only learn to construct them for toy data sets and never experience its power for group comparison tasks in real multivariate data sets.

This situation is also not favourable for extensive integrated tool use. The curriculum at the upper secondary level does not contain data analysis as such but focusses on probability and inferential statistics. Software use is encouraged and prescribed for simulation, interactive visualisations and calculations (for instance with binomial or normal probabilities). Several federal states require graphic calculators as obligatory tools in classrooms andin final and central examinations. However, current final examinations do not require an elaborate use of technology in probability and statistics beyond simple calculations.

This situation requires action on different levels. For secondary schools interested in intensive tool use, we have developed teaching material for use with Fathom (Biehler, Hofmann, Maxara, & Prömmel, 2011) and we are currently developing such material for using TinkerPlots in primary education (Frischemeier & Biehler, in preparation). We support local schools and students who work with these materials and tools in school, partly when they are writing their master and bachelor theses.

Professional development courses for teachers

In recent years, a significant change occurred in upper secondary mathematics education. Due to the new national standards (KMK, 2012), probability and statistics (which is called stochastics in Germany) have become obligatory even in final examinations such as the *Abitur*. As said above, several federal states require the use of hand-held calculators even in final examinations. These uses have created a widespread need for professional development.

Reacting to this need, we have developed material and courses for professional development courses for upper secondary teachers in the context of the German Centre for the Professional Development of Mathematics Teachers (www.dzlm.de), which have already reached some hundred teachers in several federal states (Barzel & Biehler, 2017). In these courses, we use the graphic calculators recommended in the respective states (Casio fx, TI Nspire) and Geogebra to adapt to the boundary condition these teachers have to take into account. The classroom material and professional development material builds on conceptions including the use of digital tools in stochastics teaching that we developed in pilot projects with Fathom (Biehler & Prömmel, 2010; Meyfarth, 2008; Prömmel, 2013). These ideas have been adapted to the more wide-spread used tools. As the TI Nspire's statistics and data module is based on Fathom, adaptation is more straightforward with this tool.

However, adaptation to tools that are not best suited for probability and statistics may distort original ideas. Let us briefly look at the role of simulation, which was introduced as obligatory in the national standards and state curricula. We can distinguish two educational functions of simulation: first, simulation for illustration and for making probability and randomness a part of students' experience, and second, simulation as part of the process of modelling and problem-solving. The latter requires that students learn the digital tool as a tool for simulation so that they can construct their own models and simulations. Thereby, they do not just execute pre-prepared simulations or observe simulations shown by the teacher. From our professional development courses, we learned that teachers are more willing to implement the illustrative use of simulations than the modelling use. Several factors act against the modelling use of simulation: the final examinations do not require this use from students, the time is to short to build students' simulation and modelling competence. A further factor is that the use of graphic calculators for simulations, even with the TI Nspire is much more complicated than with Fathom or TinkerPlots so that teachers tend to work with prepared worksheets and let the students do simulations themselves in most simple cases only.

5. Key attributes of software tools for leaning and doing statistics revisited: New challenges

Today, the tool situation at the secondary level of education looks to be pretty satisfactory as compared to 1997. The tools as mentioned earlier provide much more facilities than an average mathematics classroom is demanding. However, we have also to take into account which statistical and simulation methods are built into these tools. Here we find limitations that come from the limited set of methods that were implemented. Only Fathom has a set of simple hypothesis and parameter estimation procedures. Only TinkerPlots has a *loess* smoothing procedure implemented, and only Fathom supports visual comparison of distributions by displaying graphs of relative frequency, to name a few features. These features are relevant for elementary applications. So, an updated future-oriented vision of today would include the hope that in the future development of Codap all essential features and methods of TinkerPlots and Fathom will become integrated. Such a cost-free web-based tool that is available in different languages would be attractive for many classrooms.

However, it is helpful to look into the state-of-the-art of analysing tools for statistics from a broader perspective to see where future challenges are situated. Amelia McNamara resumed the analysis of requirements for software tools from both an educational perspective and from the professional perspective of working statisticians (McNamara, 2015, 2016, 2018). She takes into account recent developments in statistical practice and the emergence of data science. She distinguishes several key attributes of adequate software. We find her attributes in the first column of the following table 1 (McNamara, 2018, p. 5). A rough assessment of how well Fathom, TinkerPlots and Codap perform on these criteria is provided by the author of this text also in Table 1.

The major disadvantage of the three tools we focus on lies in the attributes 8 – 10. However, these features are indispensable concerning modern statistical practice.

Table 1 McNamara's key attributes of software, applied to Fathom, TinkerPlots and Codap

Key attributes	TinkerPlots	Fathom	Codap
1. Accessibility	Not cost-free	Generally not cost-free except for the German version	No costs
2. Easy entry for novice users	x	x	x
3. Data as a first-order persistent object	x	x	x
4. Support for a cycle of exploratory and confirmatory analysis	x	x	x
5. Flexible plot creation	x	x	x

6. Support for randomization throughout	x	x	
7. Interactivity at every level	x	x	x
8. Inherent documentation	For formula editor	For formula editor	On website
9. Simple support for narrative, publishing, and reproducibility	Using text boxes for commenting	Using text boxes for commenting	Using text boxes for commenting; publishing interactive graphs and states of analysis on the web
10. Flexibility to build extensions	Interactive worksheets in a multiple linked window system	Interactive worksheets in a multiple linked window system	Only for expert programmers

6. Software tools for leaning and doing statistics: Challenges from statistical literacy and data science education

Challenges stemming from the perspective of an updated version of statistical literacy

The first challenge comes from static and interactive infographics that can be found on the web. In the ProCivicStat project (https://iase-web.org/islp/pcs/) material for various issues of social importance was developed including the gender pay gap (Podworny, Frischemeier, & Biehler, 2018). If we use Google to find pictures associated with this phrase, we find thousands of coloured infographics that are very difficult to produce with the above tools (see Figure 1).

Many of the graphs have not a high quality from the perspective of good statistical visualisation, for instance, many of them have a high-ink-data-ratio (Wainer, 1984) and contain elements just for superficially attracting people and not for effective communication of statistical information. However, students who are used to the web may miss these features when they intend to communicate the results of their analysis. Among the graphs found, there are also many with high statistical quality that cannot be reproduced with the three tools we have been focussing on. A professional tool such as Tableau (www.tableau.com) gives an impression, which visualisation for communication of statistical information are state of the art and Fathom, TinkerPlots and Codap may look a little outdated from this perspective.

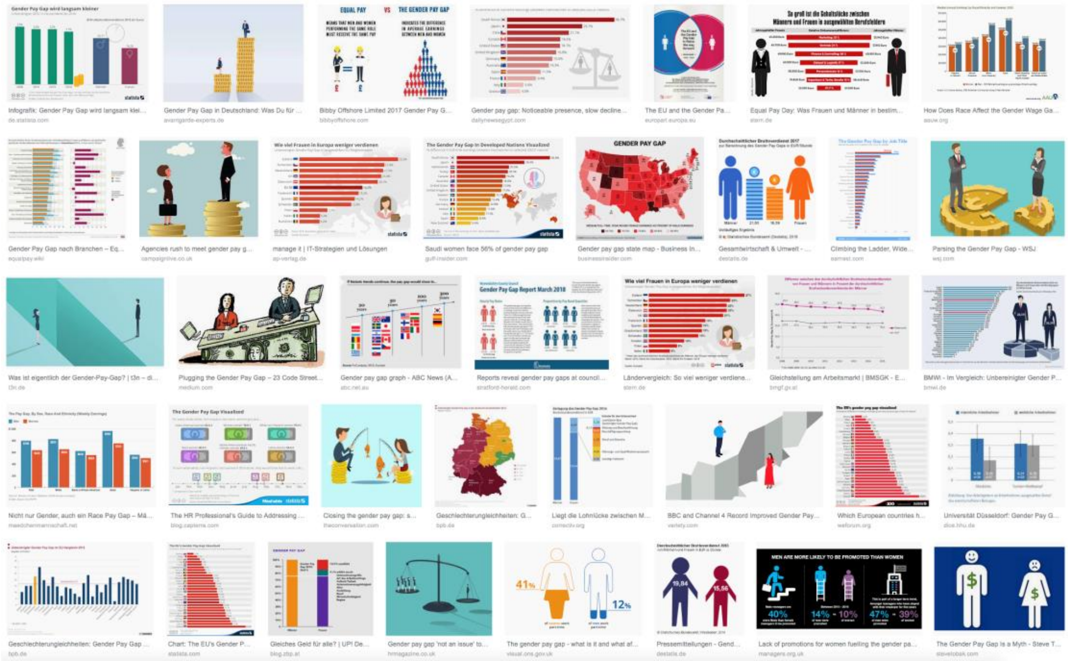

Figure 1. Results of a Google search for pictures on “gender pay gap”

A central advantage of TinkerPlots, Fathom, and Codap is that they easily deal with raw (micro) data. This feature opened the opportunity for the statistics classroom to deal with multivariate data stemming from questionnaires and other sources. Graphs and data such as depicted in Figure 1, which are essential for teaching statistical literacy, often use aggregated and summarised data. Fathom, TinkerPlots and Codap are not specialised to visualise aggregate data. Among them, only Codap easily supports the transition from non-aggregate to aggregate data employing a hierarchical data structure, which is a crucial promising step. Moreover, graphs in the web use geographical information (cities, regions, states) and visualise data in statistical maps. This feature is an important feature that only Codap is supporting.

Challenges from the emergence of data science

A related but slightly different trend that goes far beyond current conceptions of statistical literacy is captured with notions such as “big data” and “data science”. There is a world-wide change in the relationship between data and society due to the emergence of social media, data science, big data, and data-driven algorithms of artificial intelligence and machine learning, which are used in countless companies and industrial firms, often under the heading of digitalisation. In the ProDaBi project (www.prodabi.de), we are developing a yearlong experimental data science curriculum in collaboration with computer science educators for upper secondary level (Biehler et al., 2018; Biehler & Schulte, 2018). In our ProDaBi project, we are collaborating with computer science educators and include data exploration and data-driven machine learning algorithms in our curriculum.

These developments may evoke new requirements from supporting digital tools that are not fulfilled by the tools we have reviewed so far. New types of algorithms for decision trees, clustering and other types of machine learning are considered even for secondary education (Biehler et al., 2018; Gould et al., 2016). Data science also uses really big data sets the above three tools cannot cope with.

We start our curriculum with a unit on data exploration where we focus on statistical concepts and graphs and the process of data exploration and visualisation. We are using Codap as an easy entrance tool and as a result of this avoid further obstacles that would occur if we used a tool with a command language interface such as R or Python.

However, then we face the current gap between tools for learning and for doing statistics, as Amelia McNamara has been stating. We have to work on how to smooth and organise the transition to a more complex tool.

Among others, we use loudness data gathered by sensors at various places in the city of Paderborn, where each data set has about 500.000 cases, and the first tasks are "data preparation" and "data cleaning" including some averaging and smoothing, which require tool features beyond tools we focused on above. Data management has become a much important part in the practice and the education of data science than in teaching classical statistics and exploratory data analysis. The next graph shows raw data from one of the places.

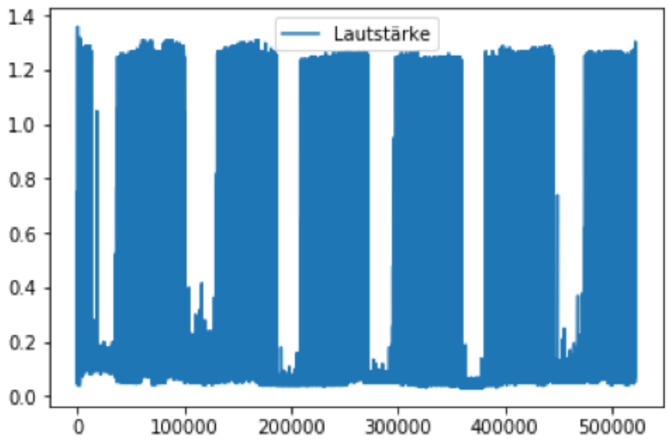

Figure 2. Graph with loudness data, collected by sensors in the city of Paderborn

For processing such data, professional tools such as R or Python are used in practice. These tools fulfil the criteria 8 to 10 (see table 1). We decided to do our classroom experiments with Python, which is closer to programming environments in computer science than R is. Python has many excellent libraries for graphical and exploratory data analysis, and algorithm libraries for machine learning (decision trees and artificial neural networks). A significant advantage is that Python can be used with Jupyter notebooks (Toomey, 2017) for an interactive report of a data analysis, where the reader can modify and re-execute code on a given data set. The code is embedded in a text with (optional) graphs that explains the analysis. Jupyter notebooks can be used as a tool for doing statistics and for communicating results of data analysis, for data synthesis, where the receiver of a notebook can interact with the analysis and the data instead of only consuming a report with pre-fabricated data analysis and no access to data. Such notebooks realise the vision of literate programming of Knuth (1984). The described features of Jupyter notebooks can additionally be used for teaching purposes. We give a brief example.

Graphs:

With the function *.plot()*, from our imported libraries, we can get Python producing a graph for us. We just have to apply the function on our data set *dfA* .

```
print(NAME.plot())
```

```
In [4]: print(dfA.plot())
```

```
AxesSubplot(0.125,0.125;0.775x0.755)
```

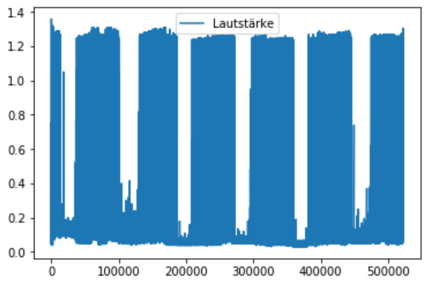

Figure 3. Jupyter notebook with Python code for preparing the loudness data

Figure 3 shows an example of a tutorial Jupyter notebook where students are introduced into using basic Python commands for data visualisation. The part of the notebook shown in Figure 4 explains to the students how to add a new variable with the time when the data were collected. Such data preparation activities are an integral part of data science, whereas traditional statistics teaching would instead work with cleaned data that have been prepared by the teacher.

Insert time and day:

We cannot see much in this graph. We need to add the information, at which day and time the data were sensored.

Until now the index is shown standing for the number of seconds that have passed since the start of the measurement.

We add a new column to the data set providing day and time.

```
NAME['Uhrzeit'] = pd.date_range('START OF SENSORING', periods =
len(Name des Dataframe), freq = 'S')
```

As *START OF SENSORING* you can input the starting time of your sensebox.

```
SenseBox A: Freitag, 06. April 2018, 9:00 Uhr
SenseBox B: Freitag, 06. April 2018, 9:00 Uhr
SenseBox C: Freitag, 04. Mai 2018, 19:30 Uhr
```

Important: Take Englisch format for the date: Englisch: mm/dd/yyyy hh:mm:ss, German: dd/mm/yyyy hh:mm:ss

```
In [5]: dfA['Uhrzeit'] = pd.date_range('05/04/2018 9:00', periods = len(dfA), f
```

Figure 4. Jupyter notebook Explaining how to add a time variable to the raw data

A significant challenge of our project is to select appropriate commands and libraries from Python or adapt its command language. We have to create a series of adequate Jupyter notebooks for supporting a smooth transition and support students in acquiring Python as an instrument for problem-solving in data exploration and visualisation on the one hand and in designing and analysing data-driven algorithms for artificial intelligence such as decision trees and artificial neural networks on the other hand.

7. Conclusions and comments

Given the experience of how much support is already needed with much simpler tools such as Fathom and TinkerPlots for organising a successful "instrumental genesis" (Monaghan, Trouche, & Borwein, 2016; Trouche, 2004) for our students, the challenge is high for statistics education and computer science education. The use of tools such as Python or R is indispensable, however, if a substantial active involvement of students in authentic data science activities is our objective. Nevertheless, the availability of a multitude of different tools for different educational purposes, rooted in a theoretical discussion on the requirements for tools for learning and for doing probability and statistics, will continue to be important.

References

Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In B. Phillips (Ed.), *Proceedings of of the Sixth International Conference on Teaching Statistics, ICOTS 6* Voorburg, The Netherlands: International Statistical Institute. Online: http://www.stat.auckland.ac.nz/~iase/publications/1/7f1_bakk.pdf.

Barzel, B., & Biehler, R. (2017). Design principles and domains of knowledge for the professionalization of teachers and facilitators - Two examples from the DZLM for upper secondary teachers. In S. Zehetmeier, B. Rösken-Winter, D. Potari, & M. Ribeiro (Eds.), *Proceedings of the Third ERME Topic Conference on Mathematics Teaching, Resources and Teacher Professional Development ETC3* (pp. 16-34). Berlin: Humboldt-Universität zu Berlin. Online: https://hal.archives-ouvertes.fr/ETC3/public/ERME_Topic_Conference_BarzelBiehler_171010.pdf

Batanero, C., Biehler, R., Engel, J., Maxara, C., & Vogel, M. (2005). *Simulation as a tool to bridge teachers' probabilistic and pedagogical knowledge*. Paper presented at the ICMI Study 15. Professional development of mathematics teachers, Aguas de Lindoia, Brazil.

Bauer, S., Biehler, R., & Rach, W. (1999). [Software] *MEDASS light Version 2.0.* Aachen und Bielefeld: Shareware.

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The new S language*. Pacific Grove: Wadsworth & Brooks.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review, 65*(2), 167-189.

Biehler, R. (2005). Strength and weaknesses in students' project work in exploratory data analysis. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 580-590). Sant Feliu de Guíxols, Spain: ERME Online: http://ermeweb.free.fr/CERME4/CERME4_WG5.pdf.

Biehler, R. (2007). Assessing students' statistical competence by means of written reports and project work. In B. Chance & B. Philipps (Eds.), *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics,* Guimaraes, Portugal, August 2007: IASE.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643-689). New York: Springer.

Biehler, R., Budde, L., Frischemeier, D., Heinemann, B., Podworny, S., Schulte, C., & Wassong, T. (Eds.). (2018). *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts*. Paderborn: Universitätsbibliothek Paderborn.

Biehler, R., & Hofmann, T. (2011). *Designing and evaluating an e-learning environment for supporting students' problem-oriented use of statistical tool software.* Paper presented at the 58th ISI Session [CD-ROM Proceedings], Dublin.

Biehler, R., Hofmann, T., Maxara, C., & Prömmel, A. (2011). *Daten und Zufall mit Fathom - Unterrichtsideen für die SI und SII mit Software-Einführung*. Braunschweig: Schroedel.

Biehler, R., & Prömmel, A. (2010). Developing students' computer-supported simulation and modelling competencies by means of carefully designed working enviroments. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics* (ICOTS8). Ljubljana, Slovenia: : International Statistical Institute.

Biehler, R., & Rach, W. (1992). *MEDASS: Explorative Datenanalyse und Stochastische Simulation - Anforderungsbeschreibung zu einem Softwarewerkzeug für Schule und Ausbildung*. Bielefeld: Universität Bielefeld, Institut für Didaktik der Mathematik.

Biehler, R., & Schulte, C. (2018). Perspectives for an interdisciplinary data science curriculum at German secondary schools. In R. Biehler, L. Budde, D. Frischemeier, B. Heinemann, S. Podworny, C. Schulte, & T. Wassong (Eds.), *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts* (pp. 2-14). Paderborn: Universitätsbibliothek Paderborn.

Efron, B. (1982). Computer intensive methods in statistics. In J. T. Oliveira & B. Epstein (Eds.), *Some recent advances in statistics* (pp. 173-181). London: Academic Press.

Frischemeier, D., & Biehler, R. (in preparation). *Datendetektiv und Datendetektivin auf Spurensuche: Bausteine zur Förderung der Datenkompetenz im Mathematikunterricht der Primarstufe*. Seelze: Friedrichverlag.

Garfield, J. B., & Burrill, G. (Eds.) (1997). *Research on the Role of Technology in Teaching and Learning Statistics. 1996 IASE Round Table Conference*. Minneapolis: IASE.

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., . . . Zanontian, L. (2016). Teaching data science to secondary students: the mobilize introduction to data science curriculum In J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education*. Berlin, Germany: IASE.

Hancock, C. (1995). The medium and the curriculum: Reflections on transparent tools and tacit mathematics. In A. A. diSessa, C. Hoyles, & R. Noss (Eds.), *Computers and Exploratory Learning* (pp. 221-240). Berlin: Springer.

Hofmann, T. (2012). *eFATHOM - Entwicklung und Evaluation einer multimedialen Lernumgebung für einen selbstständigen Einstieg in die Werkzeugsoftware FATHOM*. Wiesbaden: Springer Spektrum.

Huber, P. J. (2000). Languages for statistics and data analysis *Journal of Computational and Graphical Statistics, 9*(3), 600-620. doi:10.1080/10618600.2000.10474901

KMK. (2012). Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012). Online: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf

KMK (Ed.) (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss - Beschluss der Kultusministerkonferenz vom 4. 12. 2003*. München: Wolters Kluwer.

Knuth, D. E. (1984). Literate programming. *The Computer Journal, 27*(2), 97-111. doi:10.1093/comjnl/27.2.97

Kohorst, H. (1992). *Bevölkerungsexplosion - MEDASS Materialien*. München: FWU Institut für Film und Bild in Wissenschaft und Unterricht.

Konold, C. (1994). Understanding probability and statistical inference through resampling. In G. Cicchetelli & L. Brunelli (Eds.), *Proceedings of the First Scientific Meeting of the International Association for Statistics Education* (pp. 199-211). Perugia: IASE.

Konold, C. (2007). Designing a data analysis tool for learners. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267-291). Mahwah, NJ: Lawrence Erlbaum.

Kraft, V. (2016). Storytelling from social data: dynamic data exploration using JMP. In J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education*. Berlin: IASE.

Maxara, C., & Biehler, R. (2007). Constructing stochastic simulations with a computer tool - students' competencies and difficulties In D. Pitta & P. G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 762-771). Larnaca, Cyprus: ERME.

McNamara, A. (2015). *Bridging the gap between tools for learning and for doing statistics.* . Ph.D. Dissertation University of California, Los Angeles. Online: http://escholarship.org/uc/item/1mm9303x

McNamara, A. (2016). *On the state of computing in statistics education: tools for learning and for doing*.Online: https://arxiv.org/pdf/1610.00984.pdf

McNamara, A. (2018). Key attributes of a modern statistical computing tool. *The American Statistician*, 1-30. doi:10.1080/00031305.2018.1482784

Meyfarth, T. (2008). Die Konzeption, Durchführung und Analyse eines simulationsintensiven Einstiegs in das Kurshalbjahr Stochastik der gymnasialen Oberstufe eine explorative Entwicklungsstudie. Online; https://kobra.bibliothek.uni-kassel.de/handle/urn:nbn:de:hebis:34-2006100414792

Monaghan, J., Trouche, L., & Borwein, J. M. (2016). *Tools and mathematics - instruments for learning*. Switzerland: Springer International Publishing

Mustonen, S. (1992). *SURVO. An integrated environment for statistical computing and related areas* Helsinki: Survo Systems Ltd.

Noll, G., & Schmidt, G. (1994). *Trends und statistische Zusammenhänge. Materialien zur Explorativen Datenanalyse und Statistik in der Schule, hrsg. v.Landesinstitut für Schule und Weiterbildung, Soest*. Bönen: Verlag für Schule und Weiterbildung DruckVerlag Kettler.

Podworny, S., Frischemeier, D., & Biehler, R. (2018). Enhancing civic statistical knowledge of secondary preservice teachers for mathematics. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the*

Tenth International Conference on Teaching Statistics (ICOTS10), Kyoto, Japan: International Statistical Institute.

Portscheller, P. (1992). *Industrialisierung Deutschlands - MEDASS Materialien*. München: FWU Institut für Film und Bild in Wissenschaft und Unterricht.

Prömmel, A. (2013). *Das GESIM-Konzept - Rekonstruktion von Schülerwissen beim Einstieg in die Stochastik mit Simulationen*. Heidelberg: Springer Spektrum.

Puranen, J. (1994). Teaching data analysis with the help of Survo. In L. Brunelli & G. Cicchitelli (Eds.), *Proceedings of the First Scientific Meeting of the International Association for Statistical Education* (pp. 277-284). Perugia: IASE..

Thisted, R. A. (1986). Computing environments for data analysis. *Statistical Science, 1*(2), 259-275.

Toomey, D. (2017). *Jupyter for Data Science - Exploratory analysis, statistical modeling, machine learning, and data visualization with Jupyter*. Birmingham: Packt Publishing.

Trouche, L. (2004). Managing the complexity of human/machine interactions in computerized learning environments: Guiding students' command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning, 9*(3), 281-307. doi:10.1007/s10758-004-3468-5

Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.

Velleman, P. F. (1989). *Learning data analysis with data desk (incl. Data Desk 2.0 student version software)*. New York: W.H. Freeman.

Wainer, H. (1984). How to display data badly. *The American Statistician, 38*(2), 137-147.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.