UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Data Science Education for secondary school students and Civic Statistics Education for preservice teachers

Special lecture at JCOTS 2020
Institute of Statistical Mathematics,
Tachikawa

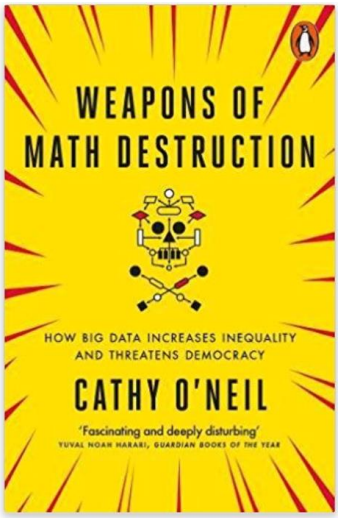
Dr. Daniel Frischemeier
University of Paderborn

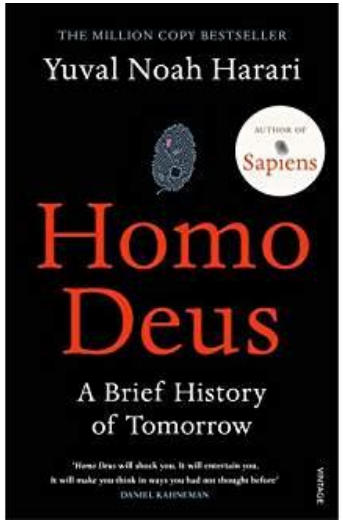
Introduction

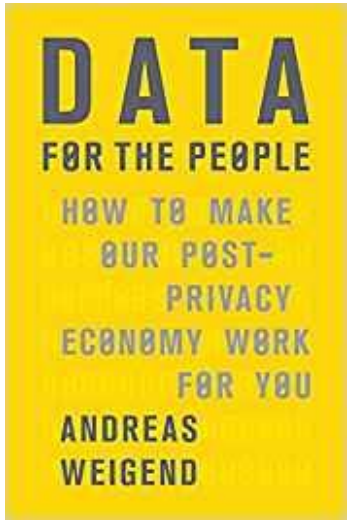
Big Data and Data Science have gained increased importance

- e.g., industrial and economic processes, marketing processes, monitoring in politics, etc. (see Gould 2017; Engel 2017)

→ Data Science permeates all areas of life and so also for students and school this field is very important



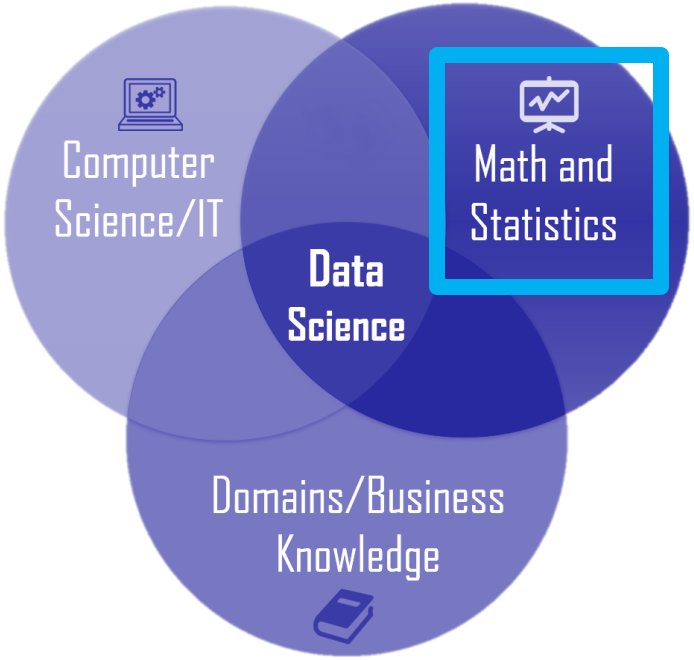




Data Science and its context

- Big Data
- Digitalisation
- Data protection in social networks
- Artificial intelligence, machine learning
- Data driven learning algorithms
- Digital humanities
- ....

- Data Literacy
- Statistical Literacy
- News Literacy
- Algorithm Literacy
- ....



And:

- **Social responsibility**
- **Civic statistics** like statistics about migration, economy, crime, health, etc.

Overview on my lecture: Two projects

Data Science Course for secondary school students

ProDaBi (**Pro**ject **Da**ta Science and **Bi**g Data in School)

- Initiated by Deutsche Telekom Foundation
- Collaborative Project of Computer Science education (WG Prof. Carsten Schulte) and statistics education (WG Prof. Rolf Biehler) at the University of Paderborn
- Aim: Developing a Data Science curriculum for secondary schools in Germany
- Further infos: www.prodabi.de

ProCivicStat (Promoting civic engagement via explorations of evidence)

- Erasmus + Project funded by the European Union
- Collaborative project (Durham, Haifa, Ludwigsburg, Paderborn, Porto, Szeged)
- ProCivicStat supports statistics teaching that enables students to engage with current social issues
- Aim: Educating teachers and developing material (tools, data sets) for teaching civic statistics in secondary school
- Further infos: http://iase-web.org/islp/pcs/

Civic Statistics Course for secondary school preservice teachers

Part 1

ProDaBi (**Pro**ject **Da**ta Science and **Bi**g Data in School)

Data Science Course for secondary school students

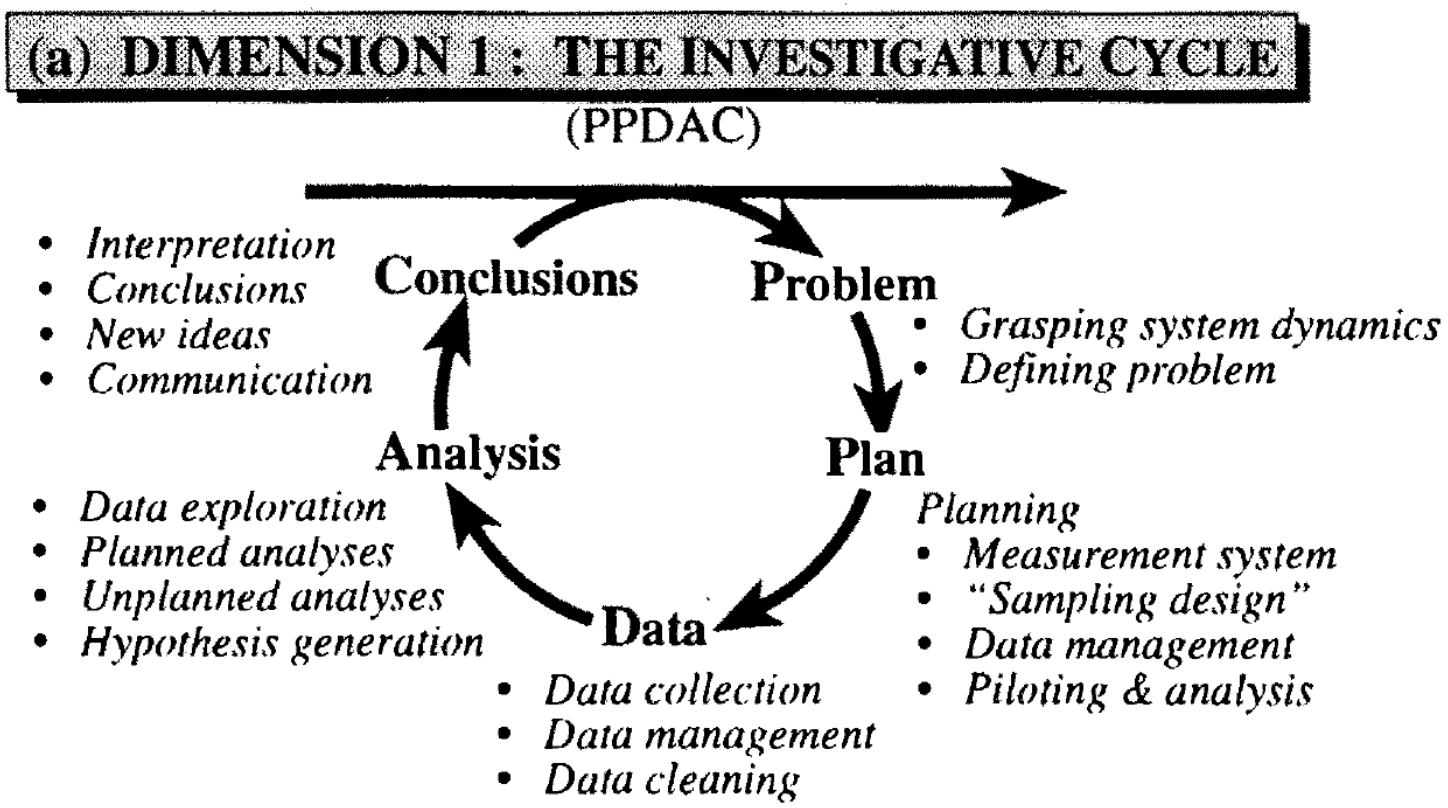
Some remarks on Data Science Education and possible implications for statistics education

Traditions in statistics education and DS

1. The PPDAC cycle of Wild and Pfannkuch (1999) & Changes in the PPDAC cycle
2. Extending the statistical view of “data”
3. Extended and new methods for data science
4. Selecting digital tools for data science: data analysis, data management, algorithm design

Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education. *International Statistical Review, 84*(3), 528–549.

The PPDAC cycle (Wild & Pfannkuch, 1999)



Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223-265. (p. 226)

Changes in the PPDAC cycle as such

- Data may be „there“ and not collected according to a plan; starting point of the cycle may be different (Tukey, 1962; Huber, 2010)
- Modelling as a step should be added
  - Classical statistical theory assumes the model to be given
  - Data science uses new types of algorithmic models (Breiman, 2001)
  - Validation of model is missing (cross validation; data for training and for testing)
  - Prediction as a goal for modelling has to be emphasized
- „Conclusions“ as a final process step has to be extended
  Statistics aims at „knowledge“; data science and computer science construct systems → „deployment“ including social responsibility

CRISP-DM cycle

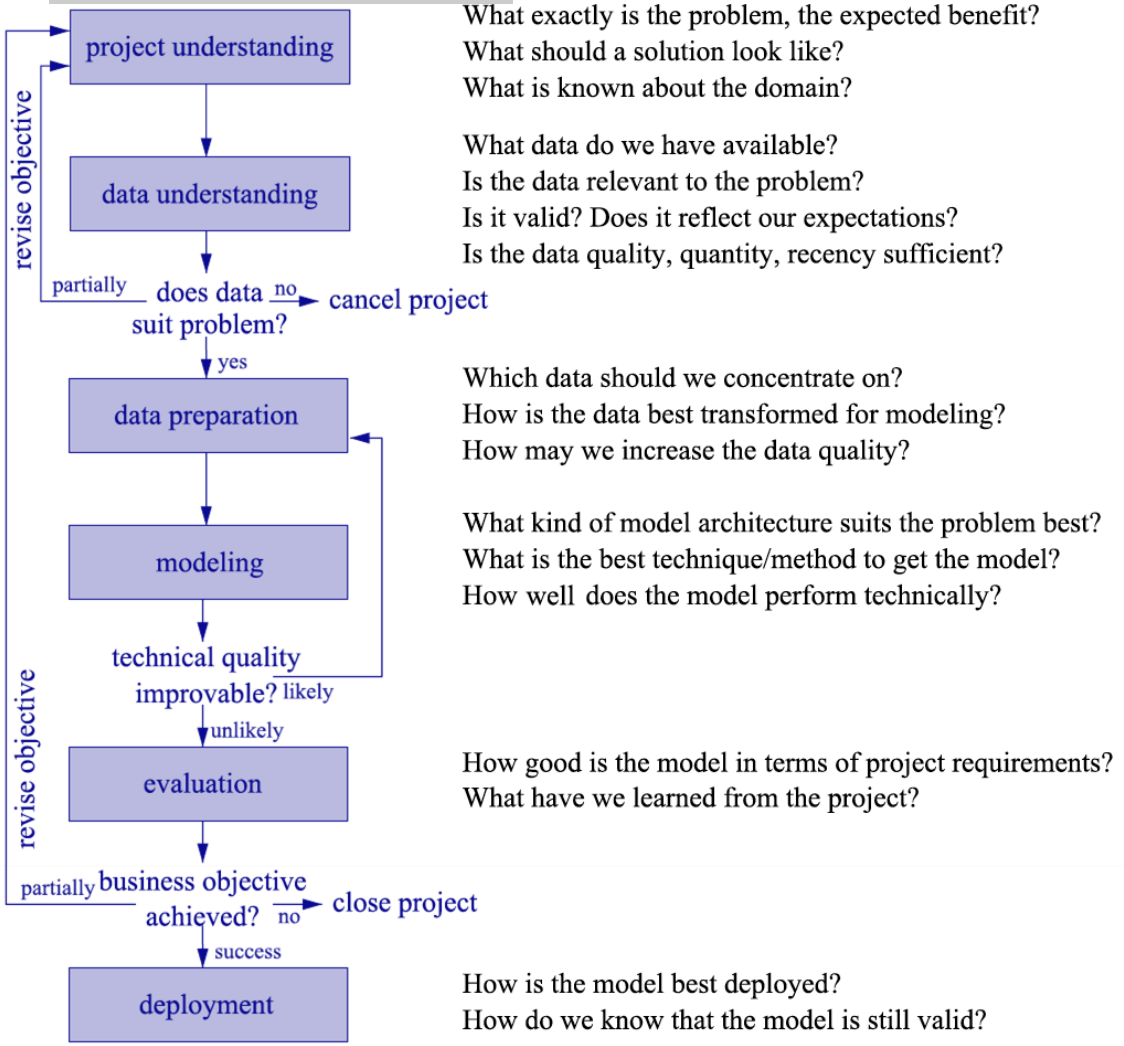


Fig. 1.1 Overview of the CRISP-DM process together with typical questions to be asked in the respective phases

Cross-Industry Standard Prozess Modell für Data Mining

See:

Berthold, M. R., Borgelt, C., Höppner, F., & Klawonn, F. (2010). *Guide to Intelligent Data Analysis - How to Intelligently Make Sense of Real Data*. London: Springer, p. 9

Extending the statistical view of “data”

- Standard in statistics (not in school):
  - The rectangular data tables with different variable types
  - Data of moderate size, multivariate data

- New types of data
  - Data collected by sensors
  - Data collected by personal devices
  - Transactional data (traffic, supermarket buys)
  - Images and texts
  - Data scraped from webpages
  - Data with geographic information

→ Big data; open data (Ridgway, 2016)

Selecting digital tools for data science

- Standard in schools (in Germany): if at all, limited use of EXCEL and GeoGebra, graphics calculator, no statistics tool

- In **experimental classrooms**: TinkerPlots and Fathom (graphical user interfaces)
  - designed for empowering students to do complex data analysis (moderate size), similar to statisticians
  - No need for „computational knowledge“
  - Algorithmic details are hidden from the user
- In **statistics**, many tools such as SPSS, R, (Python,) ...

TinkerPlots, Fathom, CODAP, etc.

GAPMINDER

CODAP

TinkerPlots®
DYNAMIC DATA EXPLORATION

python™

jupyter

→ We will have a closer look at the use of some of these tools in this lecture

Selecting digital tools for data science

- **Data science requires R or Python**
  - Supporting data management and preparation
  - Simple and advanced data analytic methods available as libraries
  - Extensibility and adaptability
  - Programming environment
- **R or Python at school level?**

  Developing a well-designed library for school use possible?

  Command interface and extensibility contributes to computational thinking?

  Programming environment supports algorithmic thinking including coding?

→ Complexity of the tool vs. concentrating on the statistical context

→ We will use Python, at least taking into account the pre-knowledge of the students (computer science courses)

Introduction

→ Implementation of these ideas in secondary school classroom

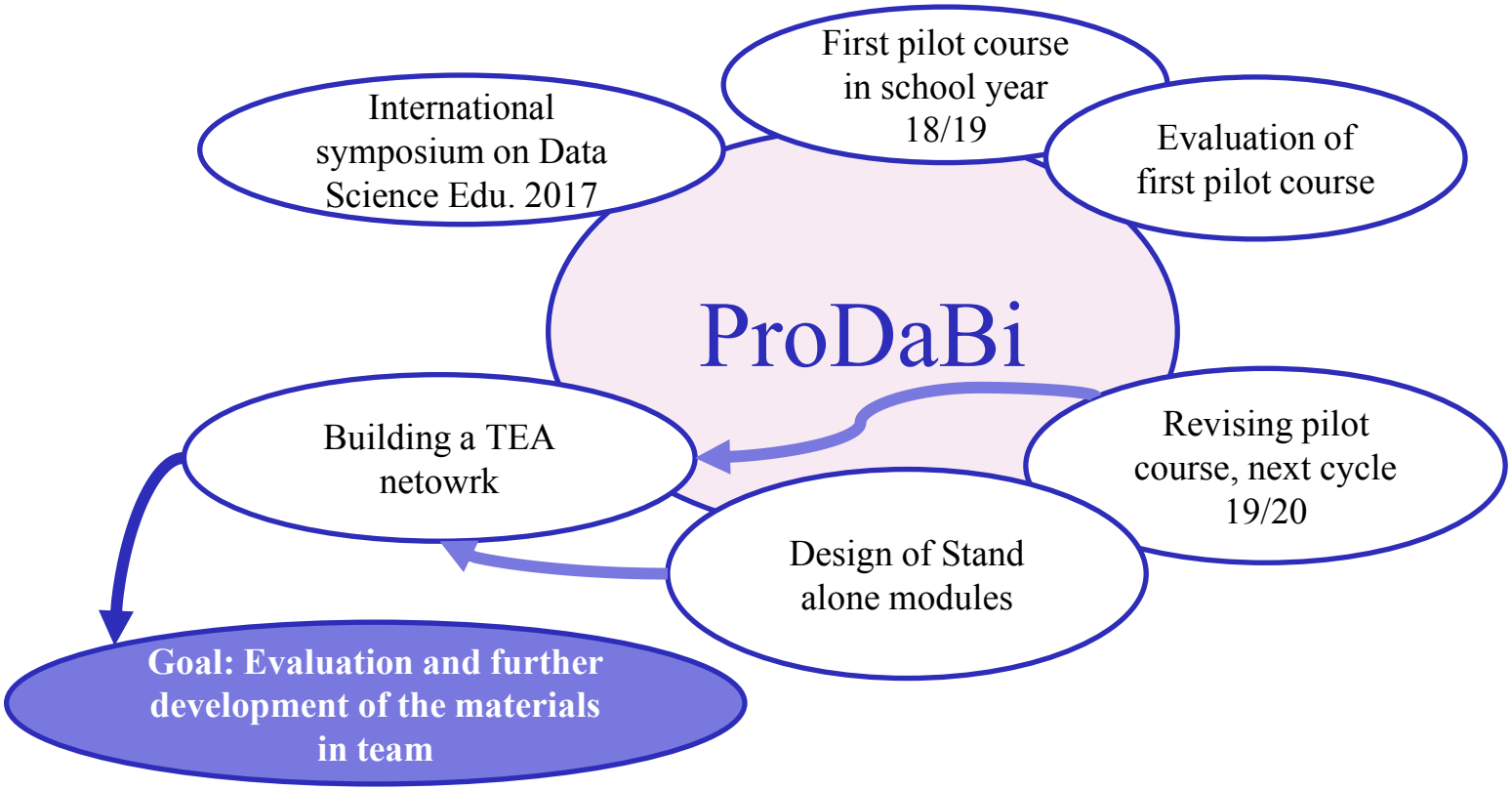
→ Project ProDaBi

The project ProDaBi

- Initiated by the Deutsche Telekom Foundation in 2018
- Cooperation of statistics education (Working group Prof. Dr. Rolf Biehler) and computer science education (Working group Prof. Dr. Carsten Schulte)
- Goal of the project: Design and Realization of a Data Science Curriculum for secondary school

  Especially: Design, Realization and Evaluation of a project course “Data Science” for secondary school

Overview: Development of the project ProDaBi



Proceedings of 2017 symposium

See:

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

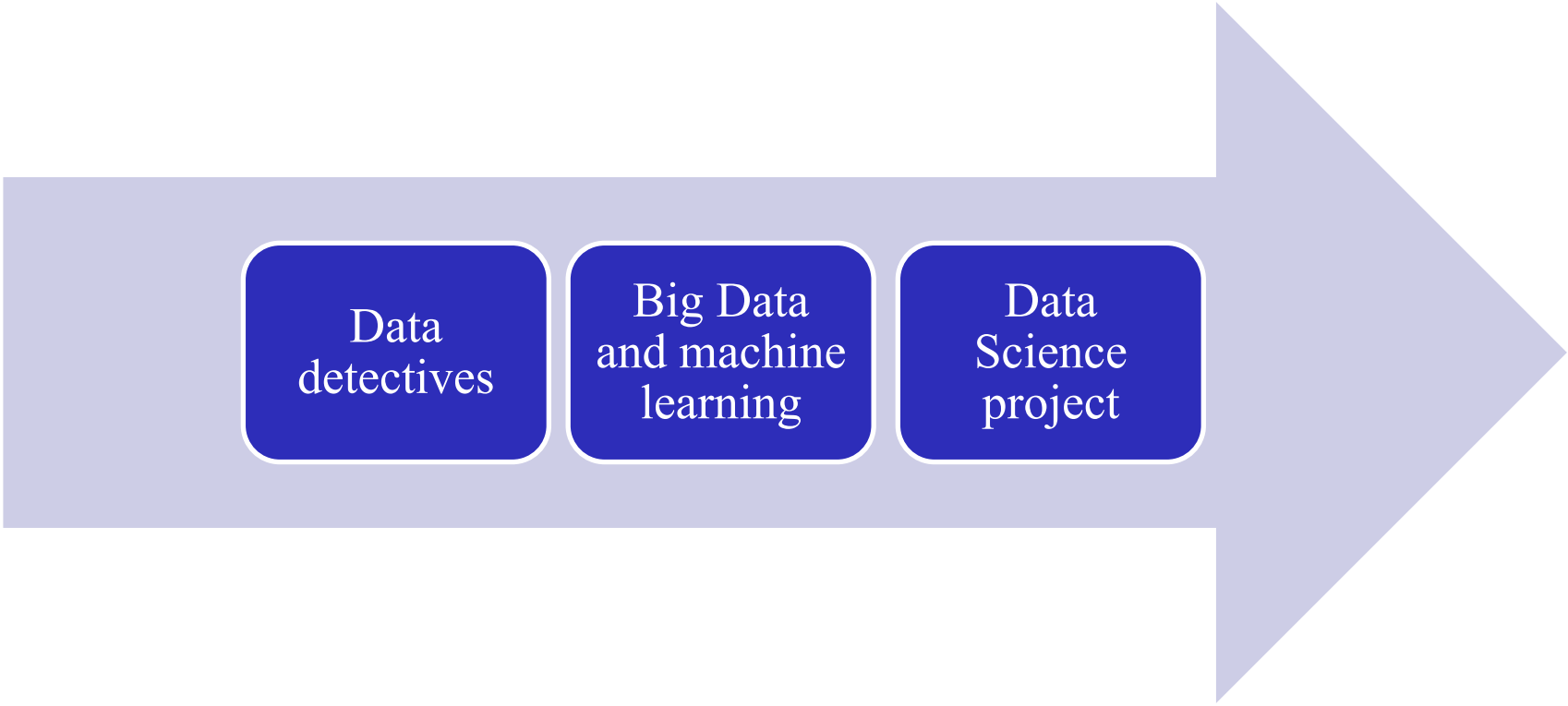
Biehler, R., & Schulte, C. (2018). Perspectives for an interdisciplinary data science curriculum at German secondary schools In: loc. cit. p. 2-14

See: http://digital.ub.uni-paderborn.de/hs/content/pageview/2931582

Experimental „Project course“

- Collaboration with two secondary schools in Paderborn
- Year-long course, 3 school hours per week
- Grade 12: students with computer science experience
- **One major teacher, 2-3 observers**
- **Material was jointly developed by the team**
- **Cycle 1: school year 2018/2019; Cycle 2: school year 2019/2020**

Pilot course in School year



Structure of the project course

- **Module 1: Data and data detectives**
  - Data exploration with CODAP
  - Data exploration with Jupyter notebooks
- **Module 2: Big Data and Machine Learning**
  - Decision trees
  - Artifical neural networks
- **Module 3: Data Science projects with partners from industry and administration**

- Our Tools
  - Codap (codap.concord.org)
  - Python; Jupyter Notebooks

The project course in the school year 2019/2020

Module 1:
Data and Data Detectives

Two units in module 1 “Data and Data Detectives”

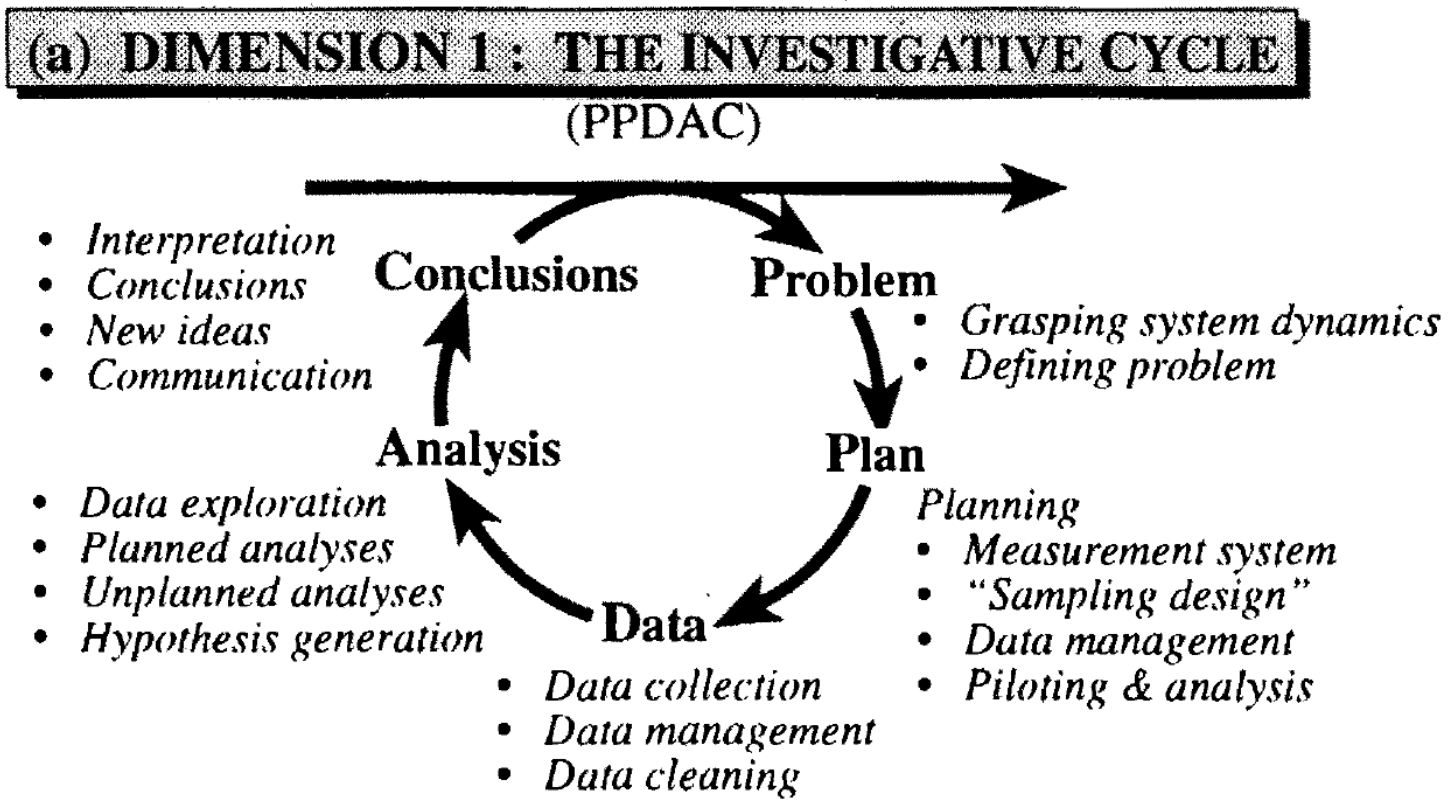
- 1. Unit: Data analysis with CODAP
- 2. Unit: Data analysis with Jupyter Notebooks

Learning goals for unit 1

Specifically we want the students in the unit 1 "Data analysis with CODAP" to…

- …explore and analyze a multivariate data set with regard to selected adequate questions,
- …use/apply basic terms of descriptive statistics and statistical concepts
- …use and evaluate digital tools like CODAP for their data exploration,
- …document and present their data analysis in an adequate form

Students were first involved in PPDAC cycle



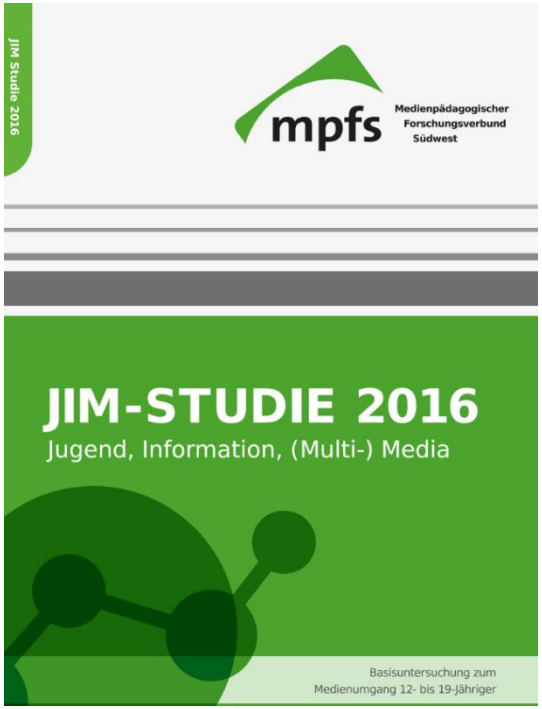
Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223-265. (p. 226)

JIM Study (Youth, Information, Media)

Representative Survey of 12 - 19 year old students in Germany

Questions:

- General information (grade, age, gender)
- Use of „classical media“ (books, journals, …)
- Use of digital media
- Use of social media, messenger, ...
- Use of Youtube
- Use of games (Computer, Tablet, console, etc.)
- ...



Online questionnaire for collecting their own JIM data

- The JIM-study includes plenty of categorized questions (e.g., "How often do you use Whatsapp?" "daily", "several times a week", "once a week", "two times a month", "once a month", "less often", "never") on the media use of German students.
- To make the JIM study, which is a nationwide study, even more interesting, we decided to modify it and to include questions which tackle numerical variables (like "How many apps to you have on your smartphone?" "How many online accounts do you have?") and we decided to collect data with the modified JIM questionnaire at the schools where our pilot course takes place.

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

JIM project

How to prepare an adequate data presentation?

Creating PowerPoint presentation
Presentation in classroom

(C) Konklusion

Generating statistical questions
What are adequate statistical questions?

Problem

Generating statistical questions alongside the JIM-dataset

Plan

Collecting JIM-PB-data in online survey

Daten

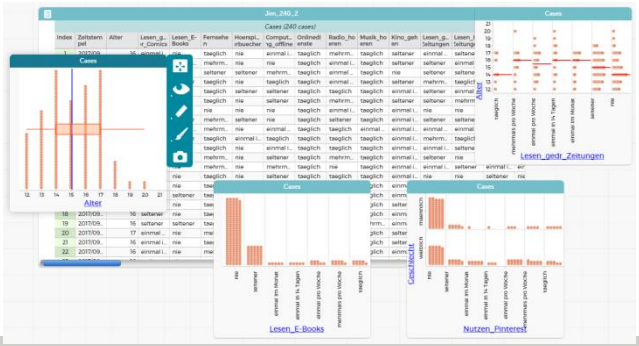
Import of JIM-PB-data in CODAP, data cleaning

Analyse

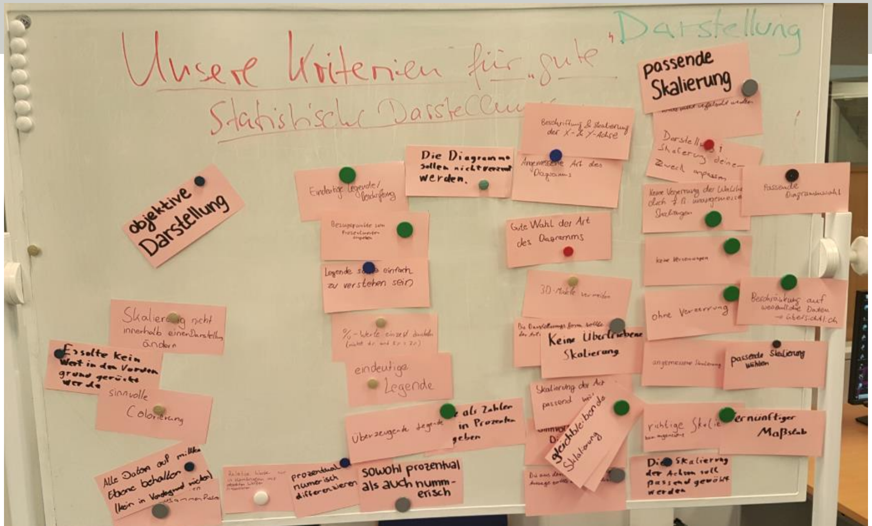
Data analysis with CODAP, Exploration of statistical question, Generating adequate diagrams

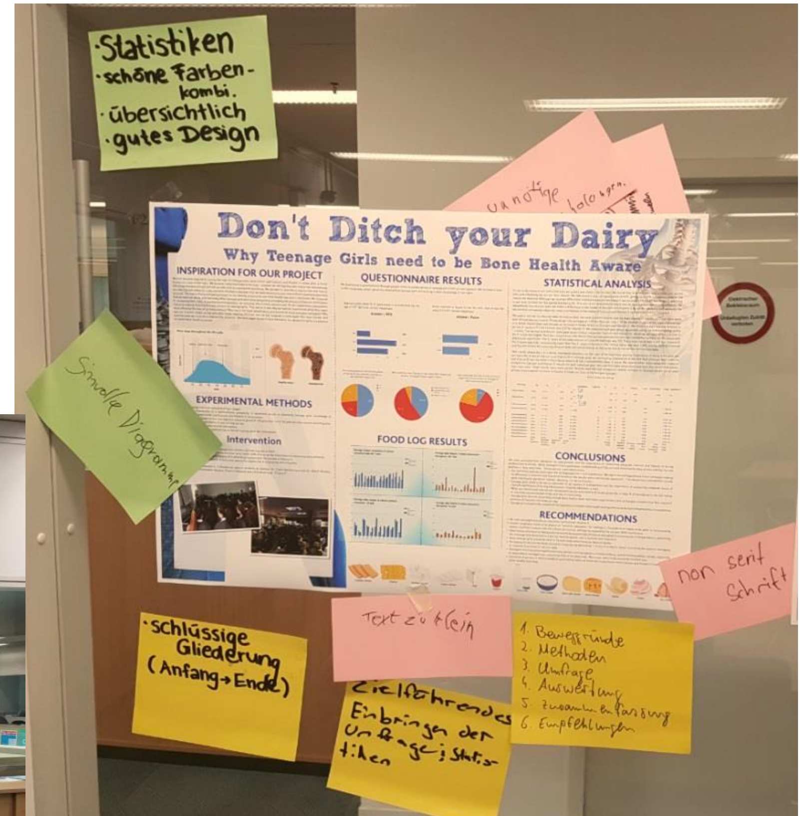
Diagrams
Percents, relationship between two categorical variables
Group comparisons

Data analysis with CODAP



UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft





Tool for data exploration: CODAP

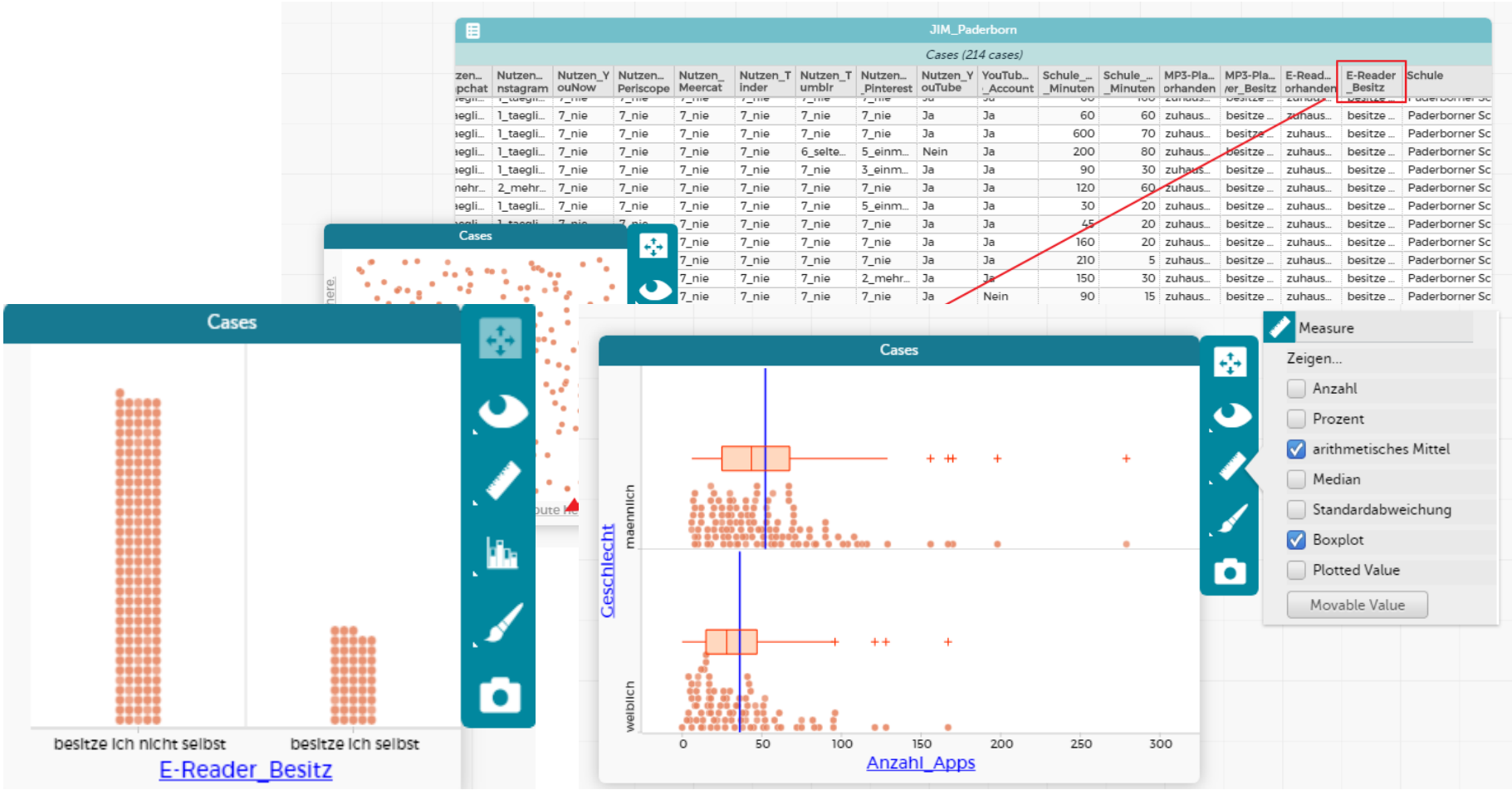
CODAP
Common Online Data Analysis Platform

a product of The Concord Consortium

- Easy entrance, quick start (with link)
- Exploration of multivariate data (interactive, graphical user interface, web-based)

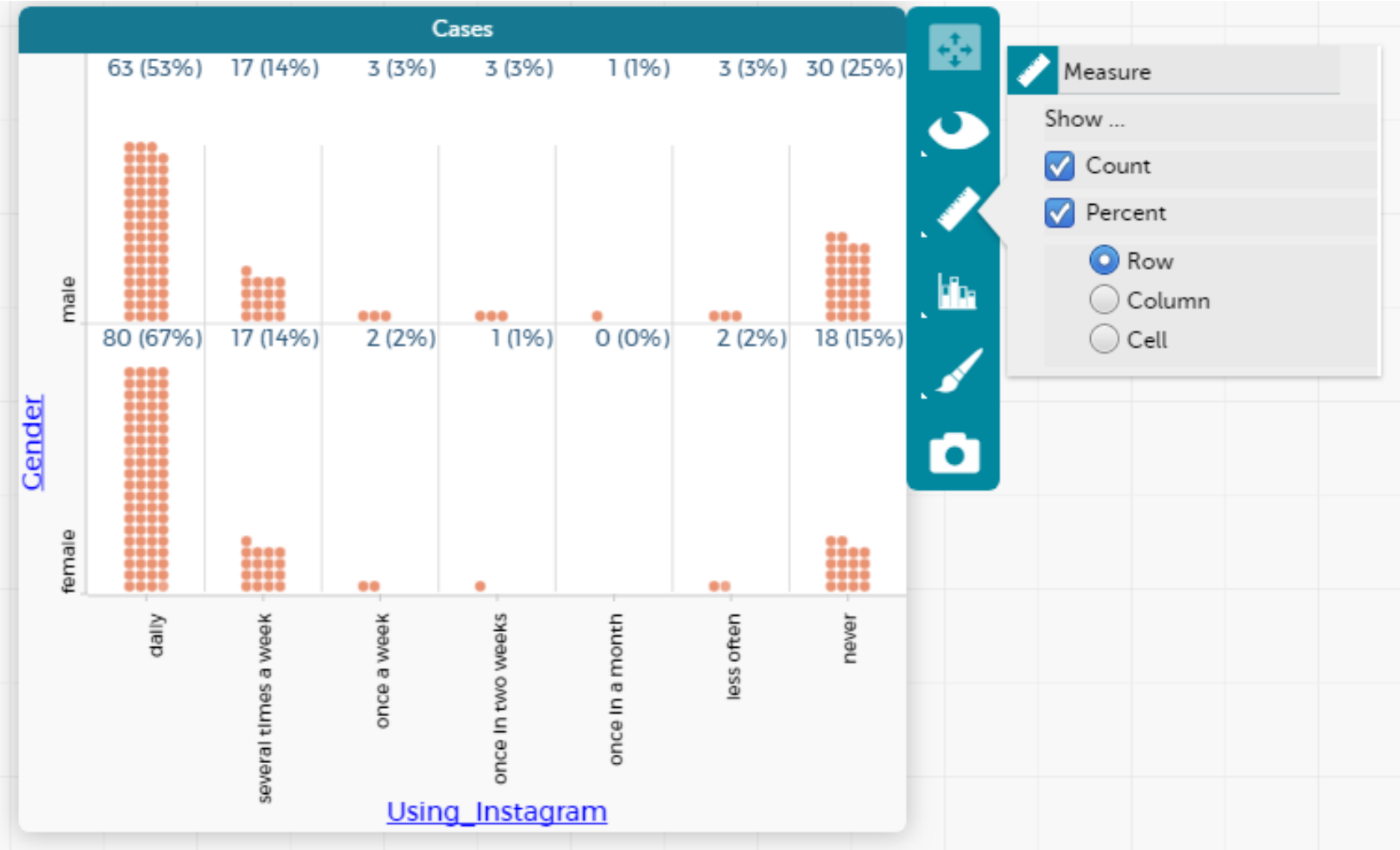
https://codap.concord.org/releases/latest/static/dg/de/cert/#shared=51283

CODAP with JIM data



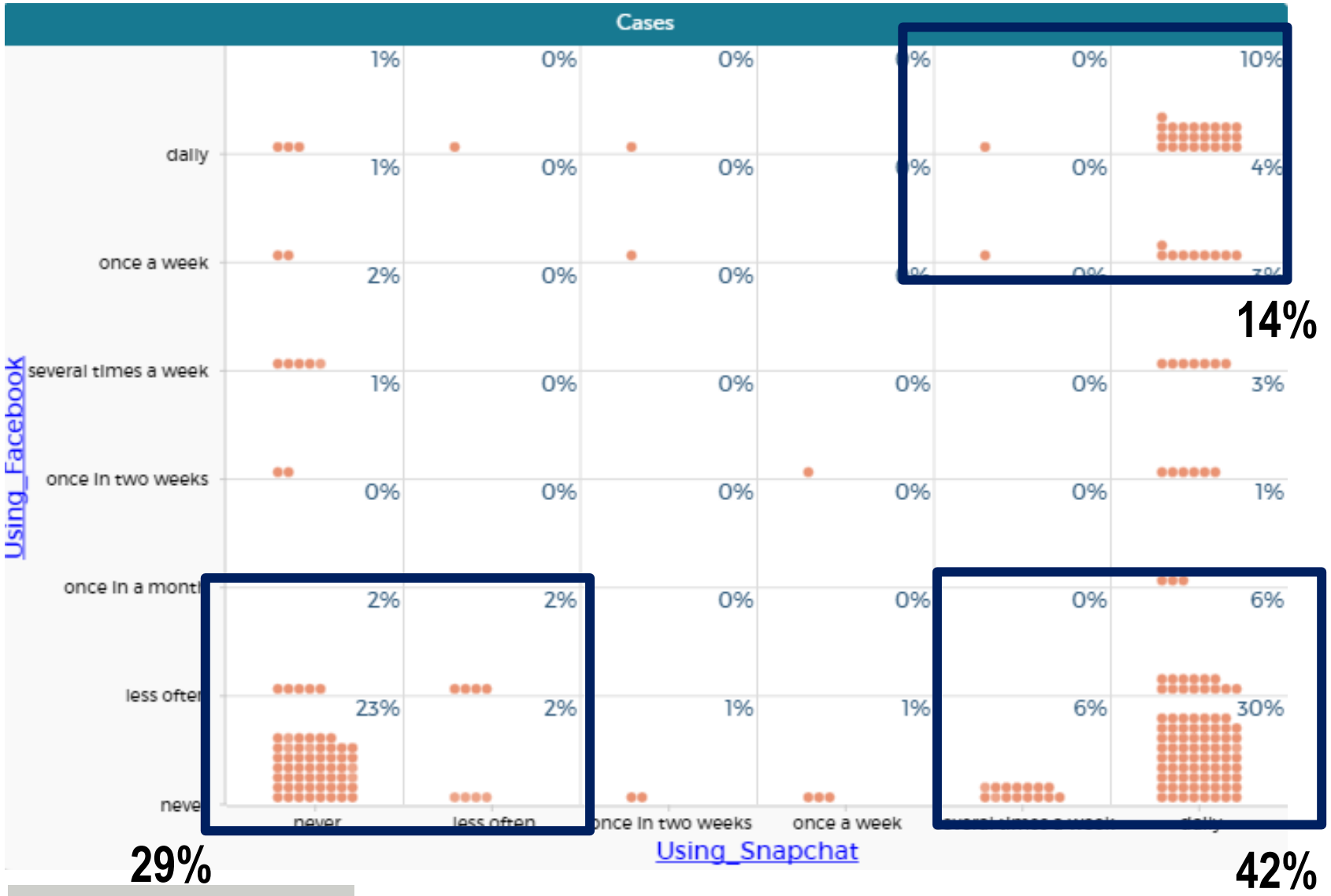
Some results of students‘ mini-projects

Instagram use: girls vs. boys



UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Different user groups: Facebook vs. Snapchat



14%

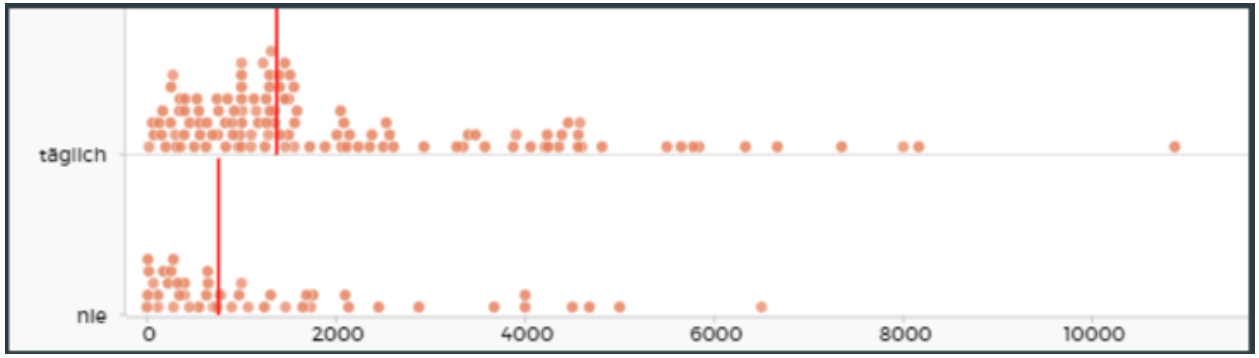
29%

42%

Frequency of Instagram use and number of pictures on one´s phone

daily

never



number of pictures on one‘s phone

Evaluation after unit 1

Item	Mean
1. The unit “Data detectives with CODAP was motivating”. (n=8)	4.25
2. I want to get more information on the topic JIM study. (n=8)	4.00
3. I was able to cope well with the demands of the unit. (n=8)	4.13
4. What I learned in the unit "Data detectives with CODAP" is important for me personally in everyday life. (n=8)	3.50
5. What I learned in the unit "Data detectives with CODAP" should become a general topic in school. (n=8)	4.15
6. The handling of CODAP was easy for me. (n=8)	4.13
7. I would like to work more often with CODAP in the future. (n=8)	3.90

Eight of the 14 students participated in the online survey and rated seven items on a **five-point** scale from "applies" to "rather does not apply"

Unit 2: Analysis of noise data:

Data analysis with Jupyter Notebooks

→ Introducing our students to a first *real* "Data Science Problem"

Data analysis with Jupyter Notebooks (Unit 2)

Jupyter Notebooks ...

- ... were originally developed to support the workflow of data analysis.
- ... are an open source tool for data science in research, education and industry.
- …consists of cells that can be individually modified and executed.

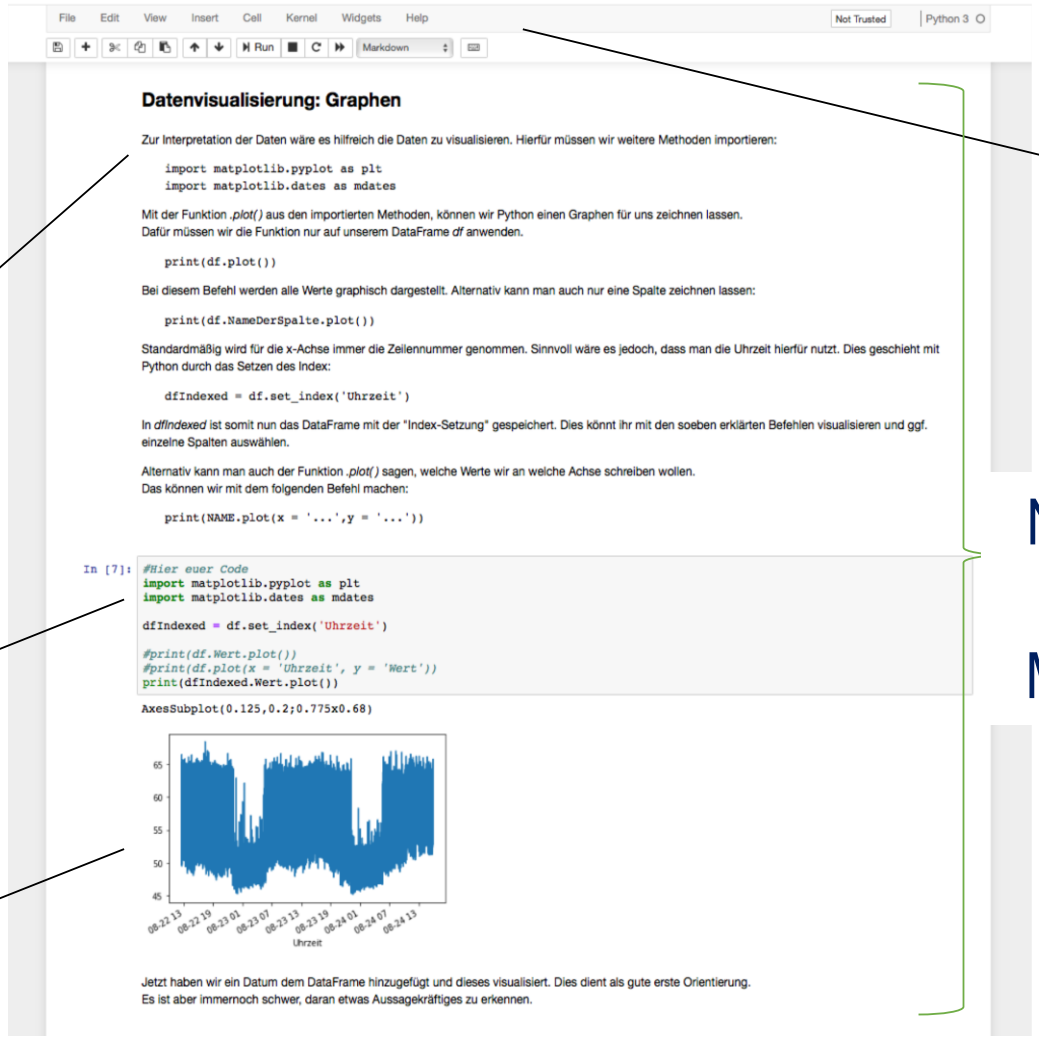
→ Differentiation between markdown cells and code cells

Jupyter Notebooks as coding environment

Markdown cell with elements like pictures, videos, etc.

Code cell

Code output



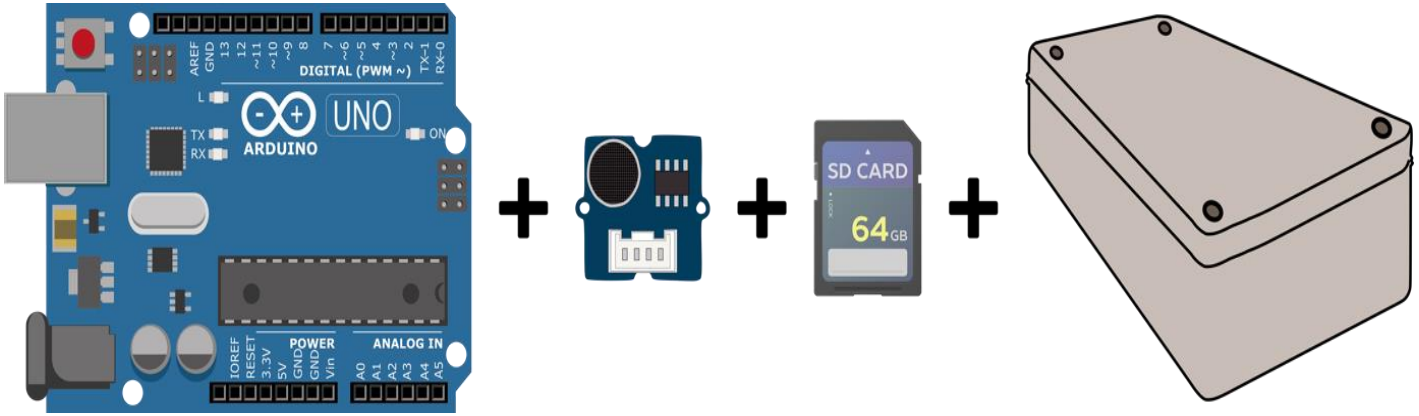
Menu and Toolbar

Notebook with Code and Markdown cell

Where is the location? Exploring noise data (Unit 2)

555.000 sensebox measurement of noise at 5 places in the city of Paderborn

Collecting data with Sense Box



Where is the location: Exploring loudness data

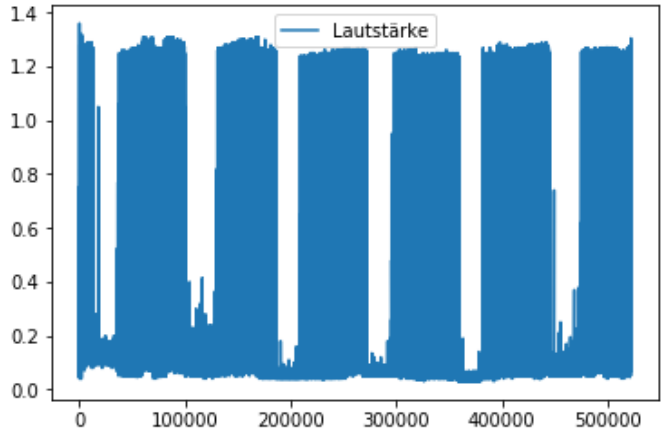
Graphs:

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)
```



Jupyter Notebooks are introduced

Environment for programming with Python

Prepared Jupyter Notebook

Pro Tag einen Graph:

Als nächsten erstellen wir einen Graph nur für genau einen Tag.

Dazu verwenden wir *.loc[ ]*

```
df_Zeitraum = df.loc[Anfang:Ende]
```

Mit diesem Befehl können wir einen Teil des DataFrames *ausschneiden*, damit wir uns diesen genauer angucken können.

Durch *.loc[Anfang:Ende]* werden aus dem DataFrame die Zeilen ausgeschnitten, welche sich zwischen der Zeile mit dem Index *Anfang* und der Zeile mit dem Index *Ende* befinden (einschließlich dieser beiden Zeilen). Wir sollten in diesem Fall das DataFrame *dfIndexed* nutzen, bei dem ihr bereits die Uhrzeit als Index gesetzt habt. Dies hat den Vorteil, dass ihr direkt die Zeitspanne mit Datum und Uhrzeit auswählen könnt.

```
df_Ausschnitt = dfIndexed.loc['yyyy-mm-dd hh:mm:ss':'yyyy-mm-dd hh:mm:ss']
```

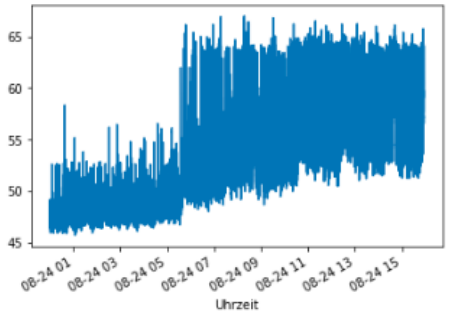
Tipp: Überlegt euch sinnvolle Variablennamen für Teile aus dem DataFrame! So könnt ihr diese später leicht nutzen, da ihr euch merken könnt, wo ihr was gespeichert habt. Eine Möglichkeit wäre, die einzelnen Tage unter ihrem jeweiligen Namen zu speichen, also df_Montag = ... ; df_Dienstag ...

.

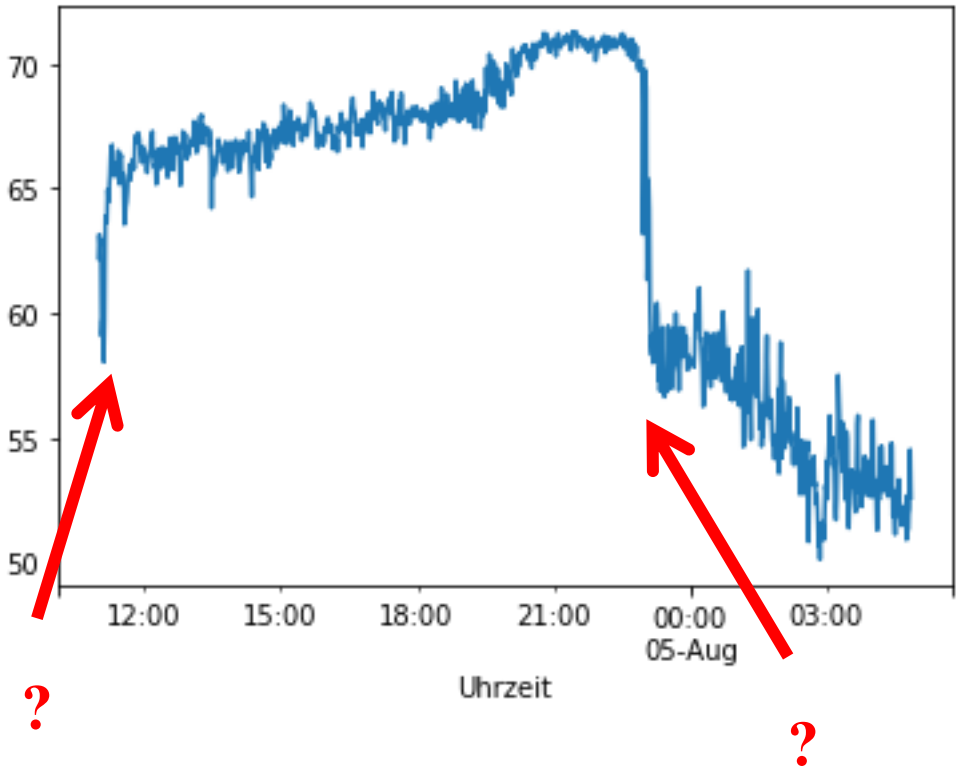
Mit print() und .plot() könnt ihr den Graphen dann wieder zeichnen lassen.

In [14]:
```
#Hier euer Code zum Auswählen!
df_Freitag = dfIndexed.loc['2018-8-24 00:00:00':'2018-8-24 23:59:59']

#Hier euer Code zum Zeichnen!
print(df_Freitag.Wert.plot())
```

```
AxesSubplot(0.125,0.2;0.775x0.68)
```



Data set 5 (4. Aug. - 5. Aug. 2018)

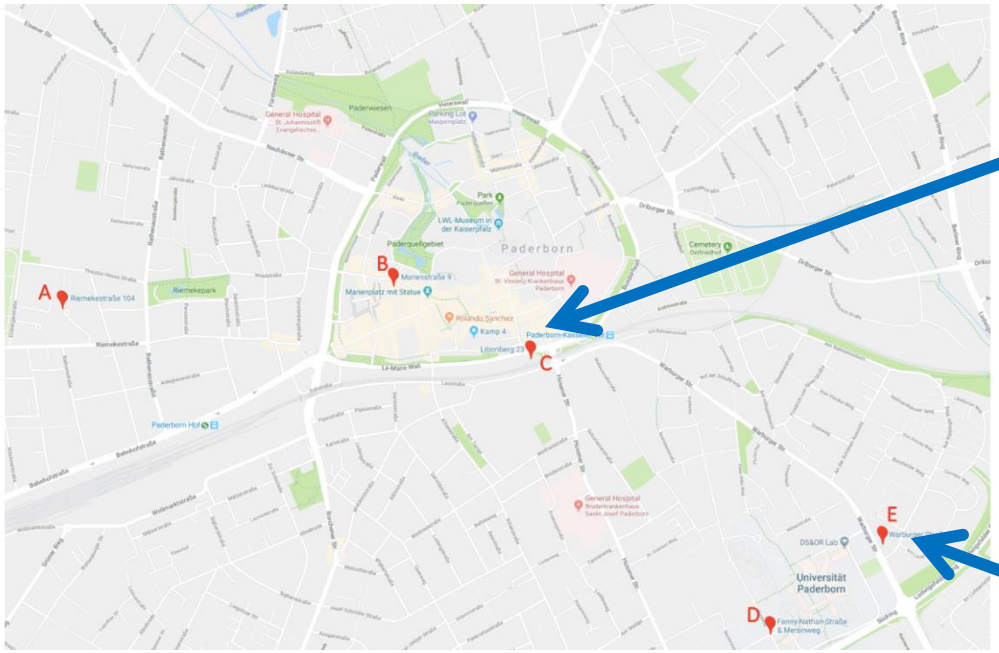


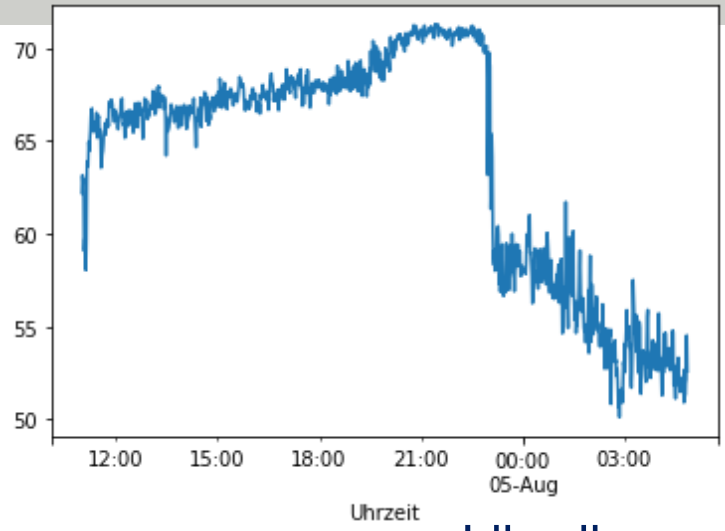
Assignment: Liboriberg

Explanation:

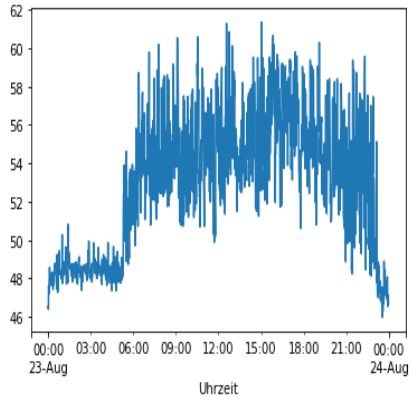
- Libori festival
- Opening 12:00
- Closing time 23:00

Allocation of further data sets





Liboriberg



Gas station

Data detectives: The noise on the track

Module 2:
Machine learning and artificial intelligence

Two units in module 2

Unit 1: Decision trees

- **Introduction into machine learning**

} 1. – 3. week

Unit 2: Artificial neural nets } 4. – 6. week

Unit 1: Decision trees

1. First introduction into decision trees

- Understanding the concept data split, Generating a tree on the basis of a dataset
- Comparison of different data splits (misclassification rate), Greedy algorithmn

2. Decision tree algorithm in form of a pseudo code

3. Programming an algorithmn

4. Applying algorithm to real datasets

- **Using decision tress to generate insights from the data**
- **Critical reflection of findings**

Unit 2: Artificial neural nets: Introduction

Artificial neural networks: Hand writing recognition with Jupyter Notebooks

Example from MNIST- data set

jupyter part2_neural_network_mnist_data Last Checkpoint: 19.07.2017 (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Run Code

In [7]:
```
# python notebook for Make Your Own Neural Network
# code for a 3-layer neural network, and code for learning the MNIST dataset
# (c) Tariq Rashid, 2016
# license is GPLv2
```

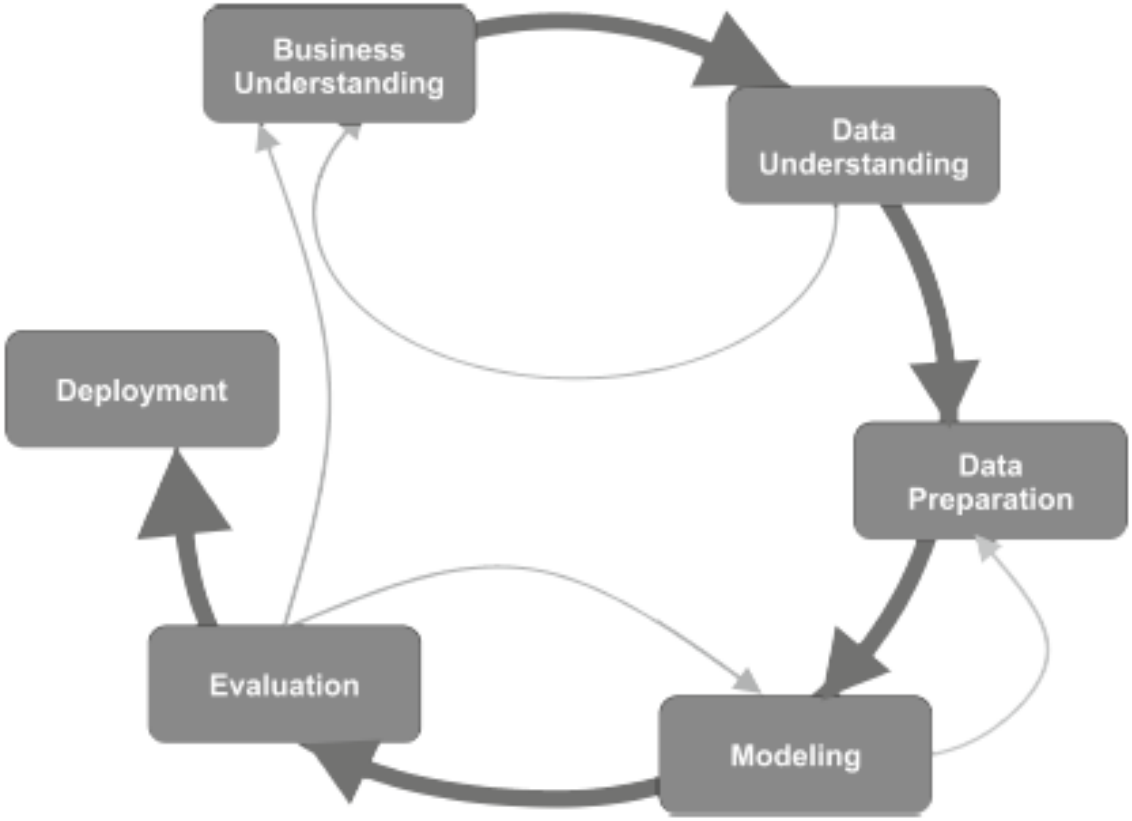
In [8]:
```
import numpy
# scipy.special for the sigmoid function expit()
import scipy.special
# library for plotting arrays
import matplotlib.pyplot
# ensure the plots are inside this notebook, not an external window
%matplotlib inline
```

In [9]:
```
# neural network class definition
class neuralNetwork:


    # initialise the neural network
    def __init__(self, inputnodes, hiddennodes, outputnodes, learningrate):
        # set number of nodes in each input, hidden, output layer
        self.inodes = inputnodes
        self.hnodes = hiddennodes
        self.onodes = outputnodes

        # link weight matrices, wih and who
        # weights inside the arrays are w_i_j, where link is from node i to node j in the next layer
        # w11 w21
        # w12 w22 etc
        self.wih = numpy.random.normal(0.0, pow(self.inodes, -0.5), (self.hnodes, self.inodes))
        self.who = numpy.random.normal(0.0, pow(self.hnodes, -0.5), (self.onodes, self.hnodes))

        # learning rate
        self.lr = learningrate

        # activation function is the sigmoid function
        self.activation_function = lambda x: scipy.special.expit(x)

        pass


    # train the neural network
    def train(self, inputs_list, targets_list):
        # convert inputs list to 2d array
        inputs = numpy.array(inputs_list, ndmin=2).T
        targets = numpy.array(targets_list, ndmin=2).T
```

Jupyter Notebook

Module 3:
Data Science Project

Modification of the PPDAC cycle: The CRISP-DM cycle



Cross-**I**ndustry **S**tandard **P**rozess Modell für **D**ata **M**ining

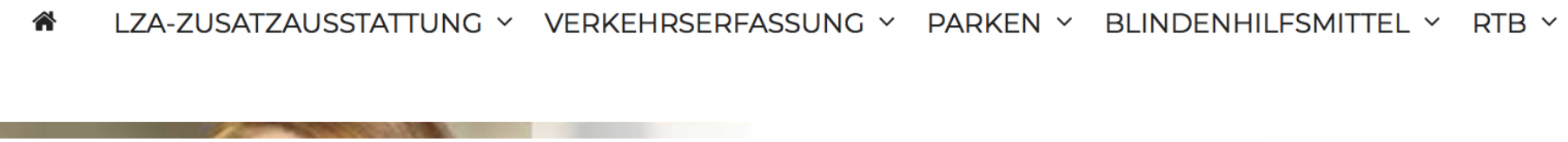
Berthold, M. R.; Borgelt, C.; Höppner, F.; Klawonn, F. (2010). *Guide to intelligent data analysis: how to intelligently make sense of real data* . London: Springer Science & Business Media.

Module 3: Data Science Project

Modul 3:

- Cooperation company RTB (Paderborn) and city of Paderborn
  - Automatic parking data recording
  - digital traffic control systems, parking spots with gates and car parks

RTB



Tasks for students

- Data group 1: ticket machine data (live) of a parking spot with gates
- Data group 2: occupancy data of a car park (live)

- Task: prediction of free parking spots

Raw data from the [illegible]or car park [illegible]gement.

Data Preparation in Python by the students (Jupyter-Notebook)

Cleaned csv-Datei of the students

	A	B	C	D
		Datum	Absolut	Relativ
0		01.01.17, 02h	80	11,30%
1		01.01.17, 03h	79	11,20%
2		01.01.17, 04h	79	11,20%
3		01.01.17, 05h	79	11,20%
4		01.01.17, 06h	79	11,10%
5		01.01.17, 07h	78	11,00%
6		01.01.17, 08h	78	11,00%
7		01.01.17, 09h	79	11,10%
8		01.01.17, 10h	78	11,00%

jupyter data_clean_up (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

```
In [ ]: #Importieren der Bibliotheken
import pandas as pd
import numpy as np

In [ ]: # Auslesen von Auslastung nach stunden

In [ ]: #Methode zum Ausschneiden der Daten defi
def cleaner(inpath,outpath):
    #Oefnen der Datei
    with open(inpath, 'r') as data :
        #Auslesen der Datei
        data_lines = data.readlines()
    data_clean = []
    #Ausschneiden der falschen Zeilen
    for line in range(len(data_lines)):
        if(line>13 and not ('Auslastung' in data_lines[line] or 'Stunde' in da
            data_clean.append(data_lines[line])
    data_reduced = []
    #Passende Spalten auswählen
    for line in data_clean:
        words = line.split(';')
        if(words[3]==''):
            data_reduced.append([words[0],words[18],words[22]])
        else:
            data_reduced.append([words[0],words[17],words[21]])
    #Dataframe erstellen
    Df_csv = pd.DataFrame(data = np.array(data_reduced),
                          index = range(len(data_reduced)),
                          columns = ['Datum','Absolut','Relativ'],copy = True)
    #Dataframe abspeichern
    Df_csv.to_csv(outpath,sep = ';')

In [ ]: #Für alle Dateien Methode aufrufen
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/01/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/02/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/03/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/04/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/05/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/06/TG/800 Auslastung/Stunden/stat806
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/07/TG/800 Auslastung/Stunden/stat800_0801
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/08/TG/800 Auslastung/Stunden/stat800_0911
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/09/TG/800 Auslastung/Stunden/stat800_1002
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/10/TG/800 Auslastung/Stunden/stat800_1102
cleaner('Daten/DatenASP/Parkhaus/TG/J2017/11/TG/800 Auslastung/Stunden/stat800_1201
```

	A	B	C	D	E	F	G	H	I	J	K	L
	Auslastung der Anlage											
	An der Talle 21											
	33102		Paderb									

linke Spalte h

Betrachtungszeitr

- Nicht Reser

Min: Minimale Zeitabschnittes (h

Max: Maximale Zeitabschnittes (h

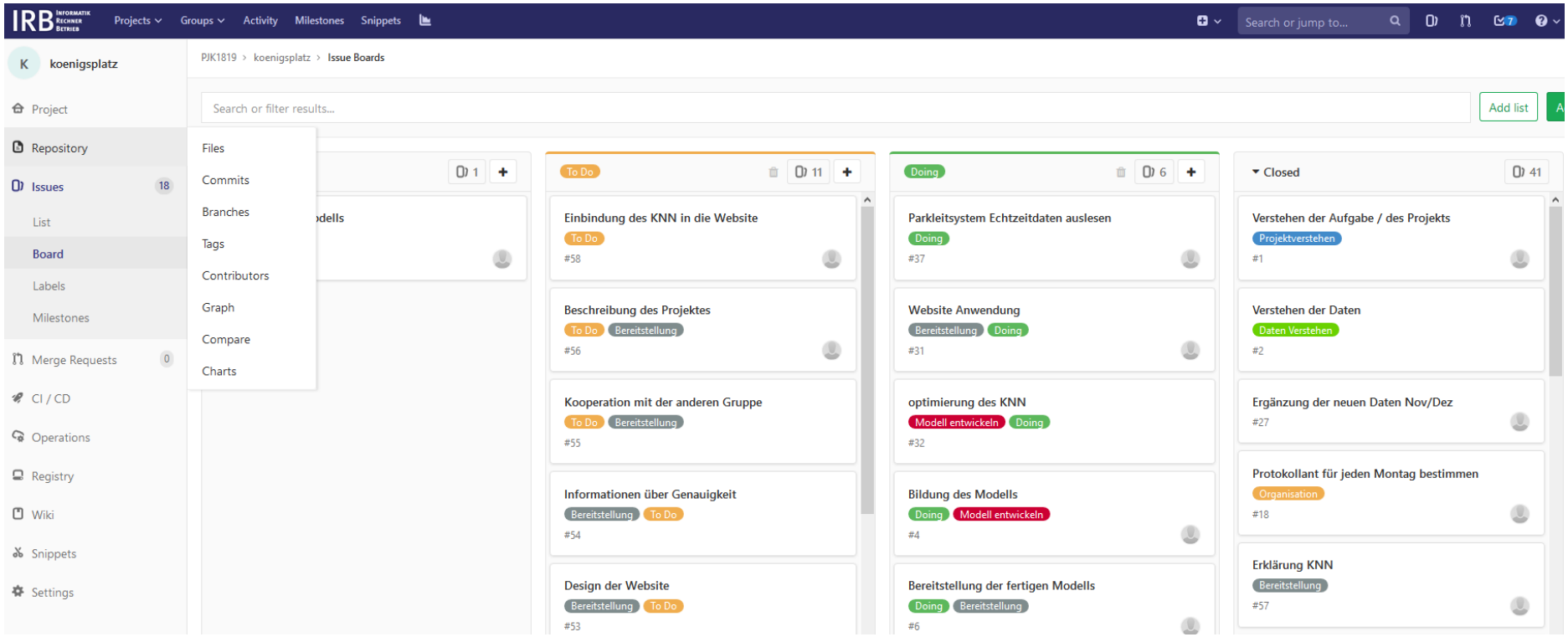
Stk.: Durchschnit Zeitabschnittes (h

Hinweis: je nach Statistikrechner Plätze für jede A Durchschnittswe Bei dem hier dar arithmetischen M

∅:

Durchschnittliche Zeitabschnittes (folgender Formel

Self organized project work organized in Gitlab



Deployment of the students

http://brainrocket.org/liboriberg/

http://www.brainrocket.org/königsplatz/

Outlook ProDaBi II

- Revised version of the project course
- Creating a network of TEAs (trusted early adopters) who will use the material
- Stepwise web-publishing of teaching material and accompanying material for „ordinary“ teachers → dissemination
- Creating 6 stand-alone units for introduction into data science, e.g.
  Data exploration with Codap

Part 2
ProCivicStat

Promoting Civic Engagement via Exploration of Evidence: Challenges for Statistics Education

Civic Statistics Course for secondary school preservice teachers

Introduction

- *Civic statistics* are statistics about key phenomena in society such as employment, health, education, social welfare, or inequality (Ridgway, 2015; Engel, Gal & Ridgway, 2016)
- Decisions in politics, society and economy are often based on c*ivic statistics*.

→Responsible citizens need statistical skills and knowledge to establish critical thinking when being confronted with statistical texts and displays in the media.

→Process of enhancing critical thinking should preferably already start in school education, but – since even teaching the traditional contents of statistics is challenging for teachers – teachers need a more focused education on civic statistics and also how to implement civic statistics in school.

The project ProCivicStat

Promoting Civic Engagement via Exploration of Evidence: Challenges for Statistics Education

The project **ProCivicStat**, funded by the ERASMUS+ program of the European Commission, aims at supporting teachers with specific courses, materials, tools, and datasets for civic statistics. (see www.procivicstat.org)

Procivicstat: 6 partner universities

PH Ludwigsburg University of Education · Durham University · U.PORTO FEP Faculdade de Economia Universidade do Porto · University of Haifa · SZTE University of Szeged · Paderborn University

Demo of materials from the ProCivicStat project

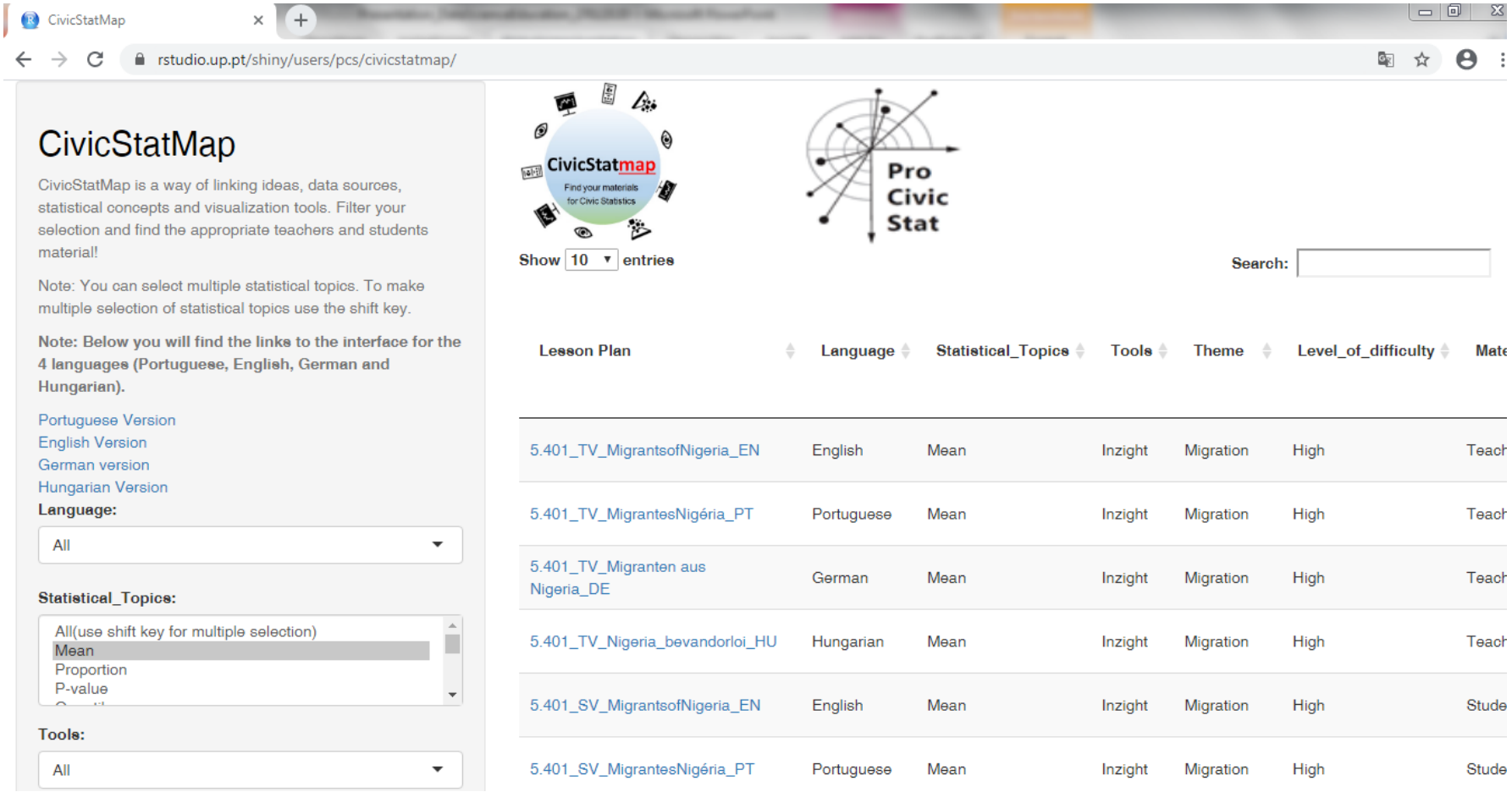
Materials from PCS (see also: http://iase-web.org/islp/pcs/)

- ***CivicStatMap → Finding materials:***
  - *Data sets*
  - *Overview on tools*
  - *Lesson plans*
  - *worksheets*

https://rstudio.up.pt/shiny/users/pcs/civicstatmap/

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Demo of materials from the ProCivicStat project



PH Ludwigsburg University of Education · Durham University · U.PORTO FEP Faculdade de Economia Universidade do Porto · University of Haifa · SZTE University of Szeged · Paderborn University

How did we implement these materials in our university teacher education courses?

Situation at our university (University of Paderborn)

Situation at the University of Paderborn for preservice teachers:

- Bachelor studies: Compulsory course on *elementary statistics and probability* and compulsory course on *didactics of statistics*
- Master studies: preservice teachers can choose a seminar to deepen and expand their knowledge they have gained in the compulsory courses

Goal:

→Design a course "Statistical literacy in mathematics classroom" embedded in the ProCivicStat project to develop the statistical content and pedagogical content knowledge of preservice teachers.

→A focus of this course is on promoting critical statistical thinking, so that future teachers have a sustainable content knowledge, pedagogical content knowledge and a positive stance towards civic statistics to bring these issues into classroom at school.

General information on the course

Situation at our university (University of Paderborn)

- First cycle of the course „Statistical literacy in mathematics classroom“ at Paderborn University in winter term 2016/2017 (Biehler, Frischemeier & Podworny, 2017)

- Redesigned course in 2017/2018
  - Seminar with 11 participants
    - Preservice teachers for mathematics in lower secondary school
    - Preservice teachers at the end of their studies, having successfully attended to a course „Elementary statistics“ and „Didactics of statistics“
  - 15 sessions (90 minutes each)
  - Idea to build on their pre-knowledge and develop statistical content and statistical pedagogical content knowledge of participants → especially in regard to critical thinking

Design of course

Learning goals

With regard to statistical content knowledge

- to deepen students´ knowledge about reading and interpreting summary statistics and graphical displays (also in the sense of *reading beyond data* of Friel, Curcio & Bright, 2001)
- to introduce students into statistical concepts and constructs (like correlation and causality, or Simpson's paradox) relevant in civic statistics
- to introduce into the definition and operationalization of concepts (like, e.g., net assets)
- to explore multivariate datasets on the base of given and self-generated statistical questions.

Learning goals

With regard to pedagogical content knowledge

- to consider contents in civic statistics across subjects
- to get to know relevant material (articles, homepages, tools, datasets, etc.)
- to learn to “simplify“ complex situations in civic statistics for use in classrooms
- to develop ideas for implementing civic statistics activities in classrooms.

Major idea of the course

Preservice teachers apply their previous gained knowledge about these statistical constructs and phenomena in more complex tasks and in project works related to specific issues of civic statistics (like poverty, unemployment, health, education, etc.).

So we want to…

- confront the participants with social phenomena in Germany, Europe and the world (e.g., gender pay gap)
- ask them to critically explore a topic taking into account investigations - starting with reports in the media, real and multivariate data from statistics bureaus
- use digital tools for data exploration.

Types of sessions - overview

Three types of sessions in our seminar:

- **Refreshing statistical and technological content knowledge**
  - Introduction to civic statistics
  - Refreshing different statistical concepts
  - Analyzing multivariate datasets with Fathom (session 2-5)
- **Students working on mini-projects**
  - Several civic statistics topics (session 6-10)
- **Gender Pay Gap (GPG) project**
  - GPG discussion about adjusted and unadjusted GPG, exploring real data about GPG and presenting results via PowerPoint (session 11-14)

Design principles of the course: SRLE

Statistical Reasoning Learning Environment (Garfield & Ben-Zvi, 2008, p. 48)

1. Focuses on developing **central statistical ideas** rather than on presenting set of tools and procedures.
2. Uses **real and motivating data sets** to engage students in making and testing conjectures.
3. Uses **classroom activities** to support the development of students' reasoning.
4. Integrates the use of appropriate **technological tools** that allow students to test their conjectures, explore and analyze data, and develop their statistical reasoning.
5. **Promotes classroom discourse** that includes statistical arguments and sustained exchanges that focus on significant statistical ideas.
6. Uses **assessment** to learn what students know and to monitor the development of their statistical learning as well as to evaluate instructional plans and progress.

Realization of the course

Sessions 1-5: Refreshing knowledge

Goal

Refresh the statistical and the technological Fathom knowledge

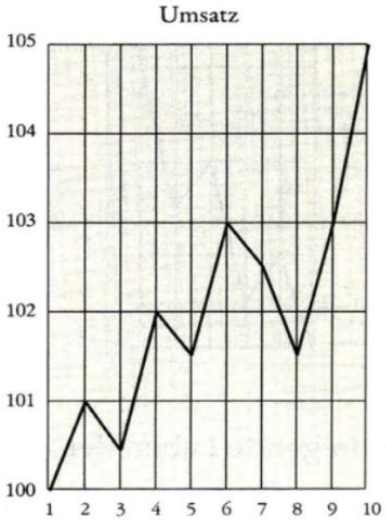
Content

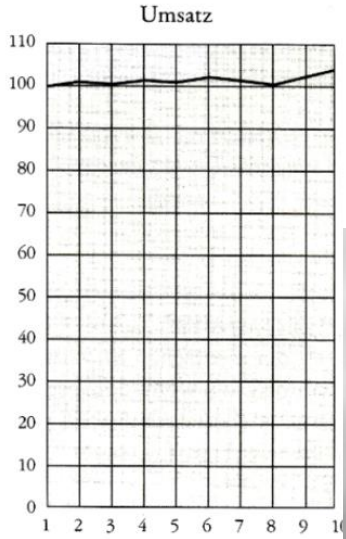
- Def. Statistical Literacy (Gal, 2002), Def. Civic statistics (Engel et al., 2016), Getting to know PPDAC-Cycle (Wild & Pfannkuch, 1999), Asking adequate statistical questions (Leavy & Frischemeier, in prep.)
- Data analysis with Fathom (analyzing student data) (Biehler, Hofmann, Maxara & Prömmel, 2011)
- percentages (row, column, cell)
- group comparisons (Frischemeier, 2017)

Sessions 1-5: Refreshing knowledge

Content (continued)

- critical thinking, how to lie with statistics (Krämer, 2007; Bauer, Gigerenzer & Krämer, 2015)





vgl. Krämer,

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Sessions 1-5: Refreshing knowledge: Statistical questions

Project: Developing statistical questions for the exploration of the dataset "Primary School NRW"

Phase 1 (Think – two persons)
Generate a statistical question for the comparison of two distributions of a numerical variable for the explo

Your statistical question (initi

_____

_____

Now present these questions

Phase 2 (Pair – four persons)
2.1 Provide Feedback

- You now have the quest course.
- The aim of this activity i partner pair. The attach
- Note the feedback belo
- Remember: The goal an improvement can only t weaknesses of the ques

Feedback:

_____

_____

**** Re-exchange *** ☺

2.2 Revise your statistical que
You have now received feedback on your questions from another couple.

- The goal is to revise your questions based on the feedback.
- Write down your revised questions.

Revised Statistical question after pair feedback:

_____

_____

Look at the question	Is it meaningful? Will the question sustain interest and curiosity of primary children? Is the intent clear and unambiguous?
Think about the variables of interest	Is the variable described clearly? Is the variable available/possible to measure?
Look at the relationship between the question and the data it will generate	Can the question be answered with a simple 'yes/no' response [avoid these type of questions]? Will the question generate quantitative data (i.e. numbers)? Will the question motivate a focus on two data sets? Does the question promote group comparison of data?
Look at (or imagine) the data	Can you answer the question with the given data? Is there sufficient data collected to answer the question? Is there sufficient variability in data collected (is there the potential for a wide range of possible data values)?

See Frischemeier & Leavy (in preparation)

Sessions 1-5: Refreshing knowledge: Statistical questions

Do girls spend more time on leisure activities per day than boys?

Initial question

To what extent do girls and boys differ in the time of their leisure activities per day?

Question after peer-feedback

To what extent do the girls and the boys differ in the time of their leisure activities in hours per day?

Question after expert-feedback

See Frischemeier & Leavy (in preparation)

Sessions 1-5: Refreshing knowledge: Percentages

Analysing Student data with Fathom

Do female students tend to get up earlier than male students?

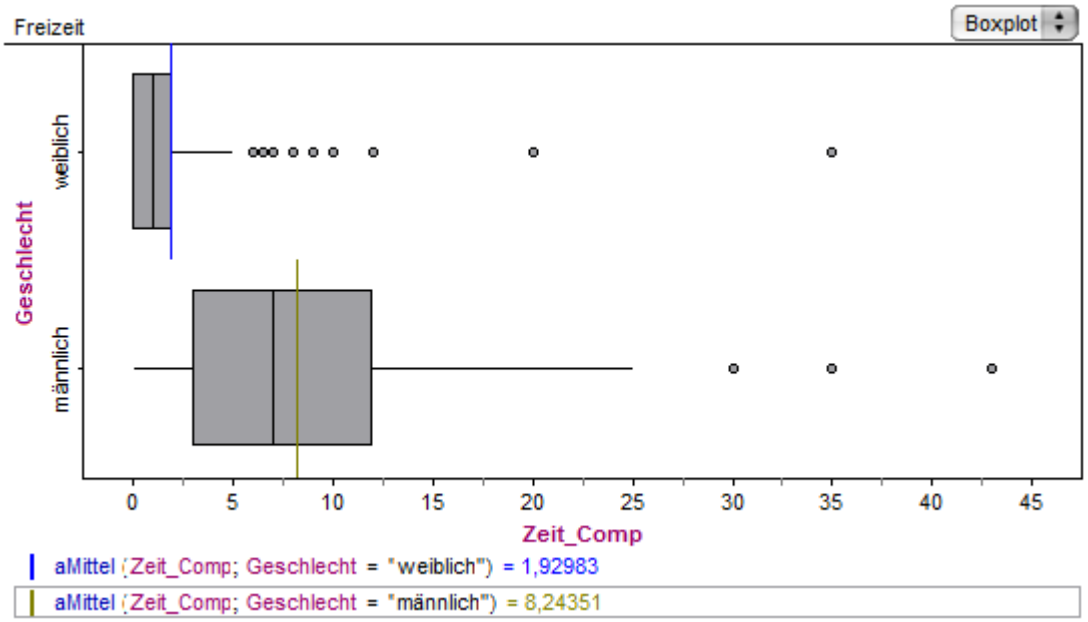
		Get_up		Zeilen-zusammenfassung
		early	late	
Gender	female	31,2	68,8	100
	male	45,5	54,5	100
Spaltenzusammenfassung		33,3	66,7	100

S1 = *Row percent*

Sessions 1-5: Group comparisons

Analysing student data with Fathom

In how far differ the female and male students in their computer use (time in hours per week)



How to compare groups?

- Center
- Spread
- Skewness
- Shift
- P-based
- Q-based

Sessions 6-10: Mini projects

- Session 6: **Statistics about the world (with CODAP)**
- Session 7: **German hospital statistics (with Fathom)**
- Session 8: **Inequality in the world (with Gapminder)**
- Session 9: **German road accidents (with a web applet from the German Statistical Office)**
- Session 10: **Daily habits of US citizens (with a web applet from the New York Times)**

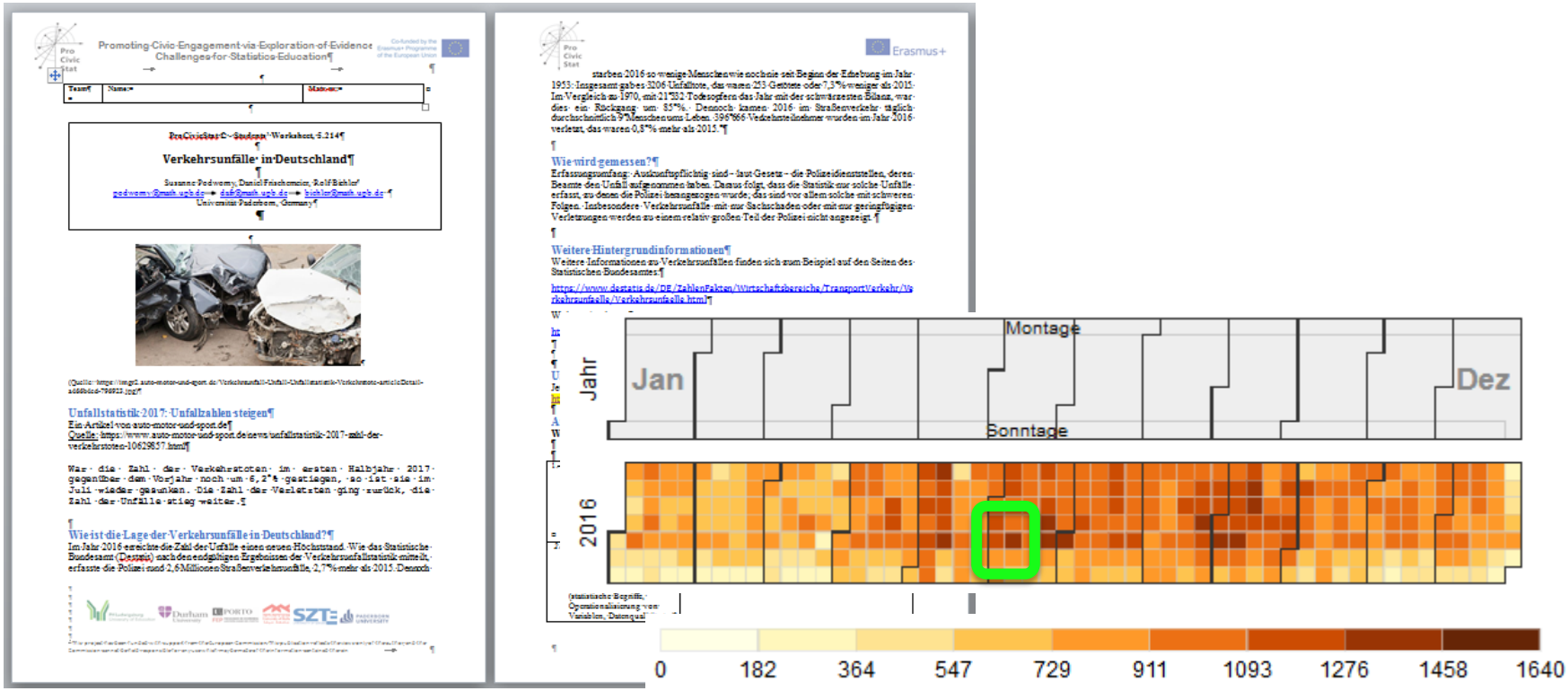
Sessions 6-10: Mini projects

Sessions 6-10 (90 mins each) were based on a worksheet

- **Introduction** (What is the topic about & why is it relevant)
- **Media report** (Newspaper article, article from the internet, …)
- **Background information & Data** (How is it measured)
- **Working phase: Analyzing data** (Pregiven and selfgenerated questions)
- **Discussing results**

Sessions 6-10: Examples

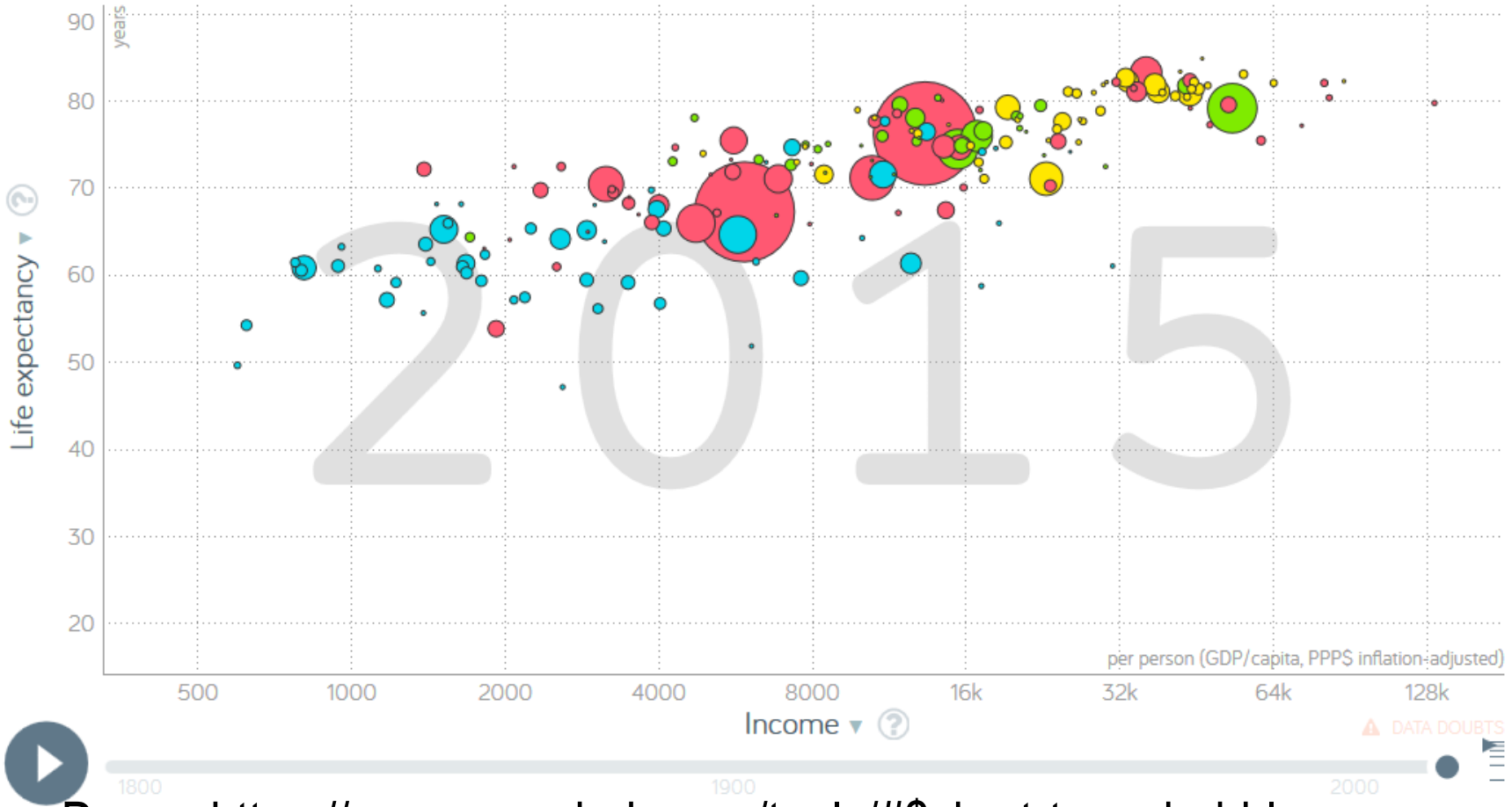
German road accidents (Destatis tool)



https://www.destatis.de/DE/Service/Verkehr/Verkehrsunfaelle.html

Sessions 6-10: Examples

Inequality in the world (Gapminder)



Demo: https://www.gapminder.org/tools/#$chart-type=bubbles

Gender Pay Gap project
(Session 11-14)

Sessions 11-14: Gender Pay Gap project

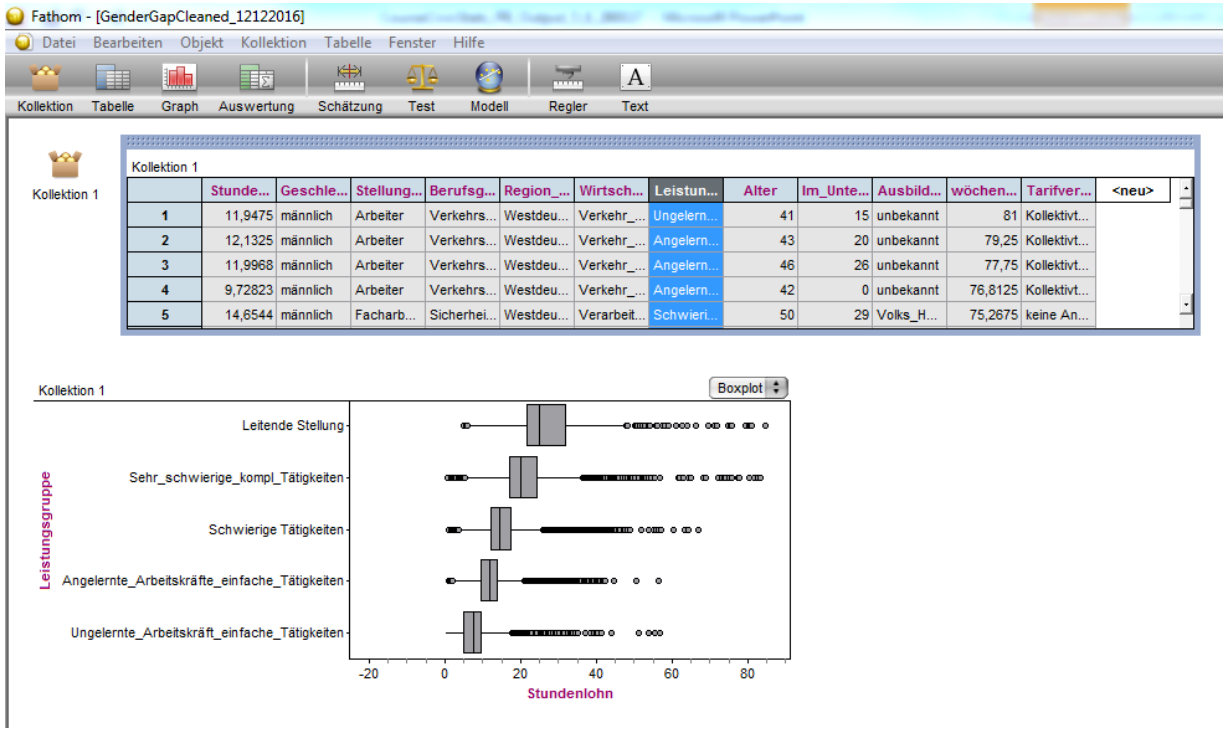
Learning goals of Gender Pay Gap project

Students should…

- get insight into some of the causes of the inequal pay situation between male and female employees in Germany,
- become familiar with the concept of gender pay gap and to be able to distinguish between the adjusted and unadjusted gender pay gap,
- explore the *German income structure* data set,
- learn to reflect media reports on gender pay gap critically and relate them to their own analyses,
- prepare their results and findings as PowerPoint presentations for their classmates

Sessions 11-14: Gender Pay Gap project

Exploring the Gender Pay Gap data in Peers with Fathom



Sessions 11-14: Gender Pay Gap project

Four sessions dedicated to Gender Pay Gap project

- Session 11 (GPG 1): Investigation on newspaper articles & background literature
- Session 12 (GPG 2): Explore data with tools in pairs
- Session 13 (GPG 3): Explore data with tools in pairs
- Session 14 (GPG 4): Present findings to colleagues (using PowerPoint presentation)

Sessions 11-14: Gender Pay Gap project, The Dataset

Dataset VSE_2006

- exported from the German Bureau of Statistics
- 60,552 cases (uncleaned) ; 59,504 cases cleaned
- employees from all levels regarding variables such as gender, wage per month, kind of employment agreement, etc.
- data set contains absolutely anonymised data for research and teaching, generated from the data “Earning Survey” (short: VSE)

Sessions 11-14: Gender Pay Gap project, several topics

Five topics (according to variables of dataset)

- Topic 1: Profession
- Topic 2: Function (simple to complicated activities, leading position, etc.)
- Topic 3: Age
- Topic 4: Economy
- Topic 5: Region

Sessions 11-14: Gender Pay Gap project, the task

Task (in peers)

Work in teams of two!

You are now to carry out a project work on the "Gender Pay Gap" with your knowledge gained in the seminar. In doing so, you should independently explore the data set for the 2006 Income Structure Survey and develop the causes for the gender pay gap on the basis of the available data.

You have learned that the differences in income between male and female workers, which are published in the media, must to be interpreted with caution because of the different aspects that determine the difference.

TASK

In this article (see link below), the focus is on the aspect “function”, which has an influence on the merit difference. Under this perspective, examine the present data set and work out the extent to which merit differences are caused by the aspect mentioned above.

Source/Link: *http://www.bild.de/ratgeber/job-karriere/gehalt-bestimmen-44537794.bild.html*

Write a short article and create a PowerPoint presentation that you will resent to your fellow students.

Sessions 11-14: Gender Pay Gap project

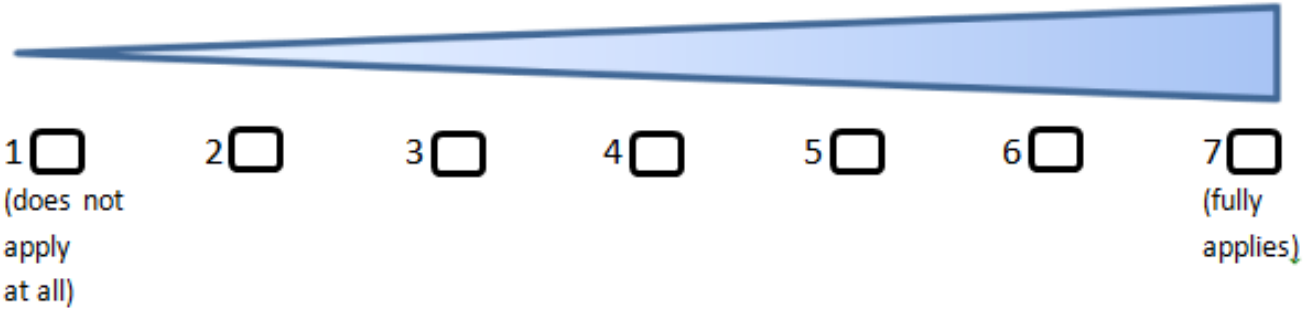
Some insights of the exploration of Gender Pay Gap data:

- Women are missing in certain professions, sectors and on the upper end of the career ladder
- Women interrupt their careers and reduce their working time for family reasons more frequently and for longer periods than men.
- Individual and collective pay negotiations have not yet succeeded in effectively overcoming the traditionally low evaluation of female-dominated professions.

Overall evaluation

Evaluation forms for sessions and tasks

1. On a scale from 1 (does not apply at all) to 7 (fully applies), rate the following statement:
   "I have understood today's task XY very well."



Please explain your rating:

2. … "I liked the contents of today's session/task very much."

3. … "The contents of today's session/task have a large didactical potential for classrooms use."

Evaluation forms for sessions and tasks

Regarding the session‘s topic/the session‘s task:

Q1 aims at cognitive aspects (understanding)

Q2 aims at dispositions (like)

Q3 aims at pedagogical aspects (pedagogical)

Evaluation: Miniprojects and Gender Pay Gap

No	Topic	Understanding Mean	Like Mean	Pedagocial Mean
6	Statistics about the world (Codap)	6	6,3	5,9
7	German hospitals (Fathom)	6,5	4,2	4
8	Inequality in the world (Gapminder)	5,8	5,2	4,4
9	German road accidents (applet)	6,6	5,6	5
10	Daily habits of US cititzens (applet	6,2	5	3,6
11-14	GPG – working	5,9	5,9	4,9

n=11

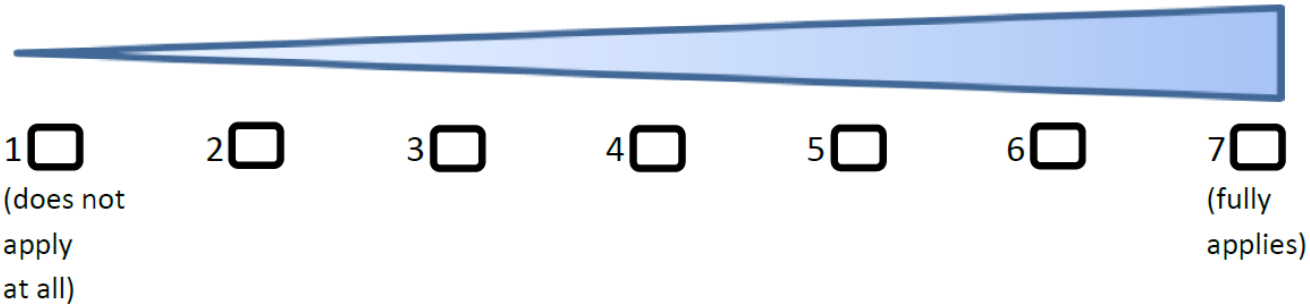
Evaluation: Miniprojects and Gender Pay Gap

No	Topic	Understanding Mean	Like Mean	Pedagocial Mean
6	**Statistics about the world (Codap)**	**6**	**6,3**	**5,9**
7	German hospitals (Fathom)	6,5	4,2	4
8	Inequality in the world (Gapminder)	5,8	5,2	4,4
9	**German road accidents (applet)**	**6,6**	**5,6**	**5**
10	Daily habits of US cititzens (applet	6,2	5	3,6
11-14	**GPG – working**	**5,9**	**5,9**	**4,9**

n=11

Evaluation form at the end of the course

1. On a scale from 1 (does not apply at all) to 7 (fully applies), rate the following statement:
   "I have understood the contents of the course 'Statistical literacy in mathematics classroom' very well."



Please explain your rating:

2. … "I liked the contents of the course 'Statistical literacy in mathematics classroom' very much."

3. … "The course 'Statistical literacy in mathematics classroom' gave me very good ideas for the promotion of statistical literacy in mathematics classroom."

Evaluation of the whole course

No	Topic	Understanding Mean	Like Mean	Pedagocial Mean
6	**Statistics about the world (Codap)**	**6**	**6,3**	**5,9**
7	German hospitals (Fathom)	6,5	4,2	4
8	Inequality in the world (Gapminder)	5,8	5,2	4,4
9	**German road accidents (applet)**	**6,6**	**5,6**	**5**
10	Daily habits of US cititzens (applet	6,2	5	3,6
11-14	**GPG – working**	**5,9**	**5,9**	**4,9**
	Whole course	**6,1**	**5,9**	**5,5**

Conclusion & Further plans

Conclusions

- Our participants worked statistically on many civic contexts, which is less common in traditional statistics courses.
- Our participants were very engaged in the mini projects and on the gender pay gap project.
- Our participants developed lesson plans for implementing some mini-projects in secondary school
- The evaluation also shows that our participants liked the exploration of German Income Structure data and the presentation of their findings via PowerPoint.

Further plans

- Further plans: Revise and repeat the redesigned course for bringing civic statistics to preservice teachers
- Design and develop further projects for secondary preservice teacher education like the gender pay gap project

Infos on both projects: ProDaBi & ProCivicStat

- **ProDaBi: www.prodabi.org**
- **ProCivicStat: *http://iase-web.org/islp/pcs/***

Questions: dafr@math.upb.de