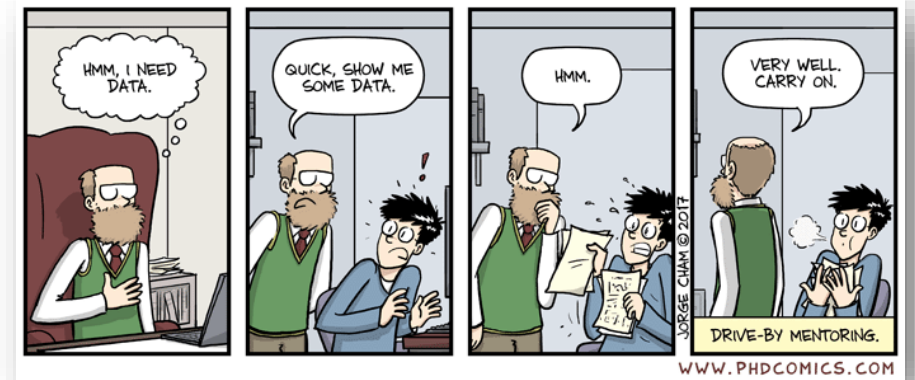


# Data Science for Human Factors

Marius Klug

Biological Psychology and Neuroergonomics  
Institute for Psychology and Ergonomics  
Berlin Institute of Technology



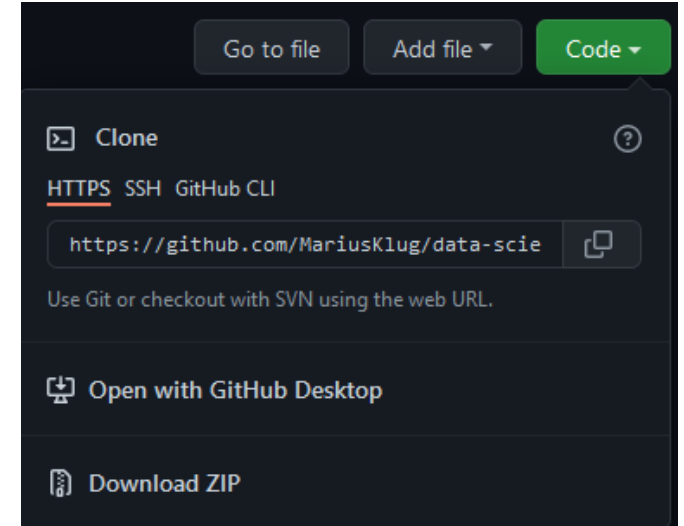
# ORGANISATION

# ISIS

- You form teams of 2 there
- You will also hand in assignments in ISIS and receive solutions automatically
- Forum for Coding Questions (help each other out!)

# Github

- <https://github.com/MariusKlug/data-science>
  - You will find all slides and code there
  - You will also find all assignments there
  - Clone to your PC or just download the zip file
- 
- It is possible that the repository gets updated. In that case I will notify you via ISIS



# General

- What you need to know:
  - General understanding of math
- What you don't need to know:
  - Programming
- What you will learn:
  - Programming with MATLAB, code structure, error messages, clean code, functions, visualization in 2D and 3D, principles of machine learning
- What will it take:
  - Blood, Sweat, and Tears. 180 hours of hard work, one does not simply learn programming!
    - Seriously, this course will probably take more time and effort than other courses, you are warned now.

# Topics

- **Introduction:** GUI and basic calculations
- **Coding 1:** Scripts, style, and variable classes
- **Coding 2:** Control statements and loops
- **Visualization 1:** Basics, subplots, get and set
- **Coding 3:** Functions
- **Visualization 2:** Descriptive plots
- **Coding 4:** Basic input and output
- **Visualization 3:** Distribution and 3D plots
- **Coding 5:** Input and output specials – last lecture before holidays
- **Machine Learning 1:** Introduction and dimension reduction
- **Machine Learning 2:** Clustering
- **Machine Learning 3:** Classification
- **Coding 6:** Efficiency and debugging basics
- **Coding 7:** Advanced functions and debugging

# Topics

- Introduction: **GUI and basic calculations**
- **Coding 1**: Scripts, style, and variable classes
- **Coding 2**: Control statements and loops
- **Visualization 1**: Basics, subplots, get and set
- **Coding 3**: Functions
- **Visualization 2**: Descriptive plots
- **Coding 4**: Basic input and output
- **Visualization 3**: Distribution and 3D plots
- **Coding 5**: Input and output specials – last lecture before holidays
- **Machine Learning 1**: Introduction and dimension reduction
- **Machine Learning 2**: Clustering
- **Machine Learning 3**: Classification
- **Coding 6**: Efficiency and debugging basics
- **Coding 7**: Advanced functions and debugging

# Not Topics

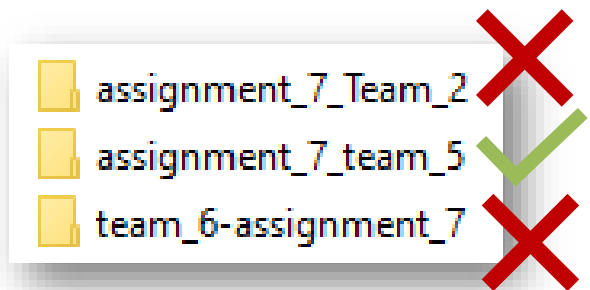
- Statistics
- Regression
- Toolboxes like Tensorflow
- Deep learning (artificial neural networks)
- Text or language processing





# Structure

- Integrated Lecture, all online on **Zoom**!
  - Zoom **Meeting ID** will be **on ISIS**
  - Discussion of **questions** about previous topics
  - New **slides** with topic overview
  - **Live coding** with new concepts
  - Explanation of **new assignment**
- Coding exercises as assignments for teams of 2 each week (**14\*4.5=63P**)
  - Teams in ISIS
  - Filename: „assignment\_<assignmentnumber>\_team\_<teamnr>.zip”
  - **WRONG FILENAMES WILL BE IGNORED! (no joke)**
  - No names as comments in the assignment, just team number!
- Large assignment (alone!) at the end of the semester (**40P**)
- Sign in for the course in QUISPOS!



# Grading

Im Wahlpflichtbereich müssen unbenotete Module im Umfang von mindestens 6 und höchstens 18 LP belegt werden. Alle Module zum Erwerb von Basiswissen und -fertigkeiten und von grundlagenorientiertem Vertiefungswissen werden alternativ benotet oder unbenotet angeboten. Die Studierenden entscheiden bei der ersten Anmeldung zur Modulprüfung, in welcher Bewertungsart sie das Modul ablegen. Die Entscheidung ist unwiderruflich und gilt auch für etwaige Wiederholungsprüfungen.

# Grading

- Either with or without grades
  - Sign in in QUISPOS **yourself!**
  - Find a partner with the same grading decision!
  - Tell me!
- General
  - Pass/fail, each exercise will just be **evaluated whether or not the code runs** and produces the correct result
  - 50% of the points needed to pass the class
    - This means **80% in weekly assignments is enough to pass** without the final assignment
- Special
  - Depending on the situation, code **MIGHT** also get points for partially correct exercises, and actual feedback about the code might be provided

Notenschlüssel:	
Mind. Punktzahl	Note
95:	1,0
90:	1,3
85:	1,7
80:	2,0
75:	2,3
70:	2,7
65:	3,0
60:	3,3
55:	3,7
50:	4,0
<50:	5,0

# MATLAB

- TU Berlin offers **MATLAB licences** for free
  - <https://de.mathworks.com/academia/tah-portal/tu-berlin-31461245.html>
  - All my code runs on **2019b**, so best get that version to avoid problems!
- Toolboxes:
  - Statistics and Machine Learning
  - Signal Processing

# Why MATLAB?

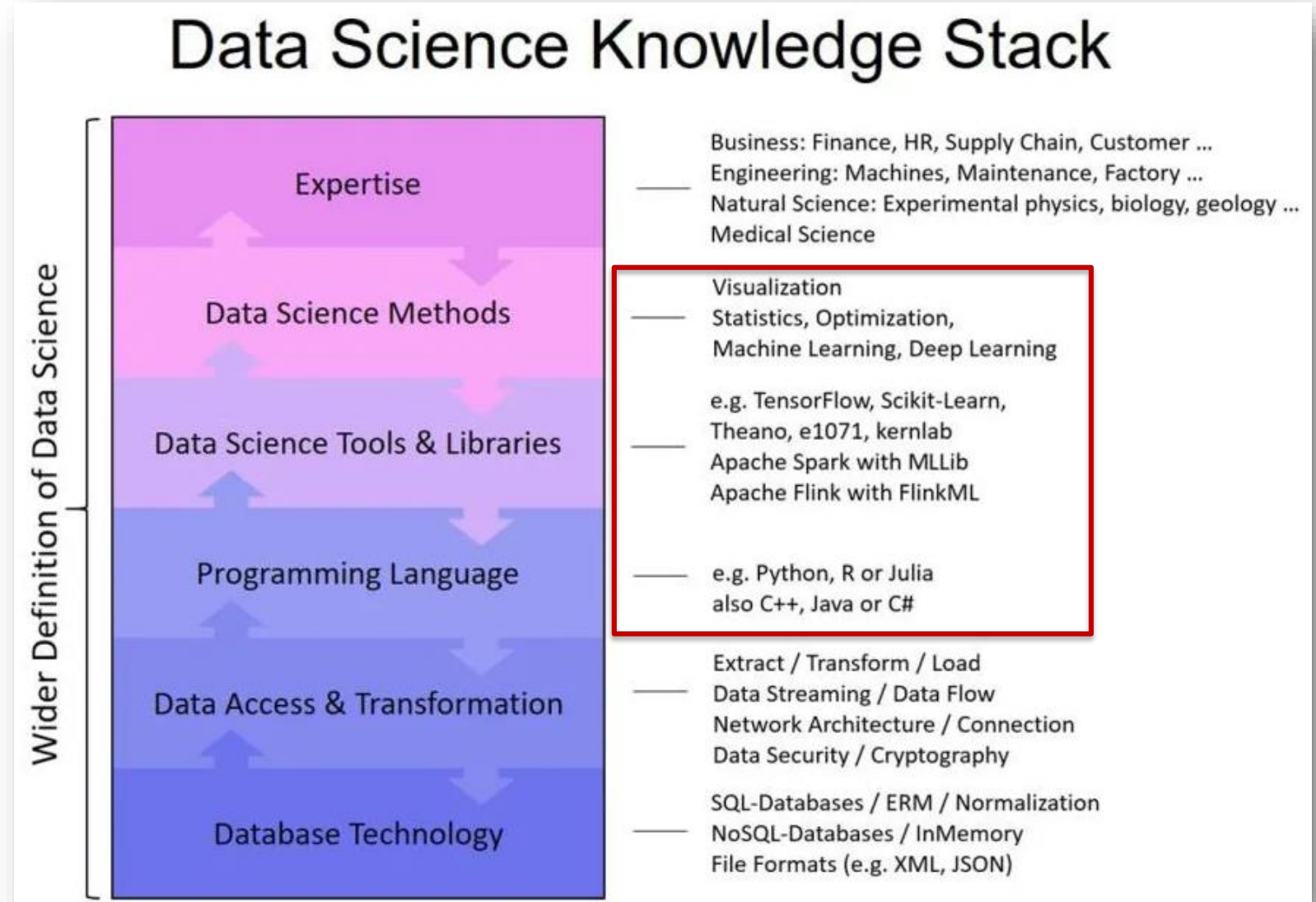
- Easy to interact directly with data
- Easy to learn, forgiving, high-level language
  - Errors won't directly make your PC blow up
- Widely used in academia, also our lab
- Comprehensive documentation and great debugging tools
- Vast online resources (StackOverflow, FileExchange)
- I know MATLAB best ;)
  - Skills are transferable anyways
- Alternatives are e.g. Octave (very similar), R, Python, Julia, C++

The GUI, MATLAB as a fancy calculator

# INTRODUCTION

# Data Science

- Gain knowledge from data
- Generate recommendations for actions
- Visualize data
- Data Mining more specifically about trends and patterns
- Programming is essential

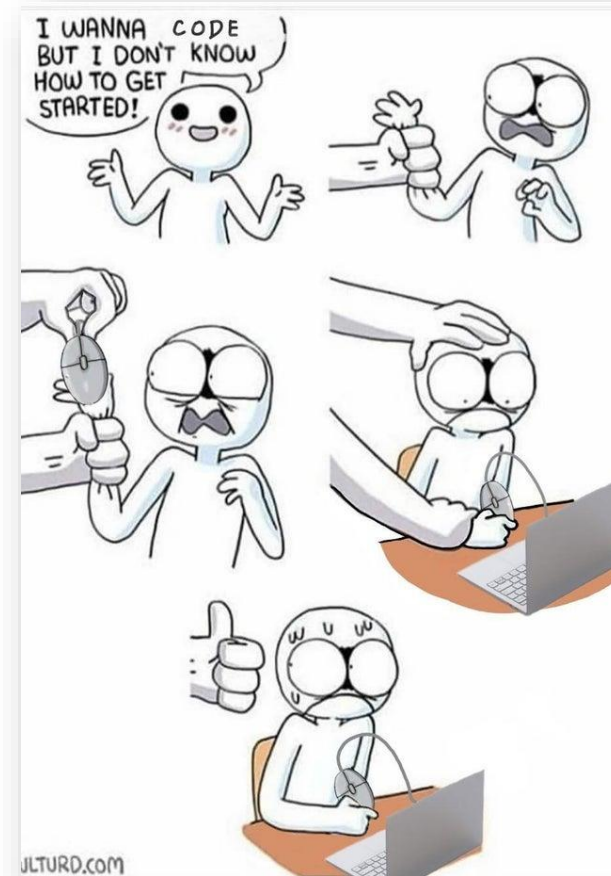


<https://www.cio.de/a/was-ein-data-scientist-wirklich-koennen-muss,3577657>



# Become a programmer

- Learn the words
- Learn the grammar
- **Speak (Code)**
- Fail and learn
- **Learn to tolerate frustration**
  - Don't use wireless mice.
  - Have some anger management.
  - Seriously.
  - I'm not joking.



**ALWAYS CODE  
AS IF THE GUY  
WHO ENDS UP  
MAINTAINING,  
OR TESTING  
YOUR CODE  
WILL BE A  
VIOLENT  
PSYCHOPATH  
WHO KNOWS  
WHERE YOU  
LIVE.**

*~dave carhart*



# MATLAB

The screenshot displays the MATLAB R2016b interface with the following components:

- Top Toolbar:** Includes tabs for HOME, PLOTS, APPS, SHORTCUTS, EDITOR, PUBLISH, and VIEW. Below these are icons for New, Open, Find Files, Import Data, Save Workspace, New Variable, Open Variable, Analyze Code, Run and Time, Preferences, Set Path, Add-Ons, Help, and Community.
- Current Folder:** Shows the file structure of the current project, including folders like 'nback\_matlab' and 'actual\_runtime.m'.
- Editor:** Displays the script 'beatiful\_plots.m' with the following code:
 

```

1 figure(3); clf
2
3 % generate data
4
5 % rng default % for reproducibility
6 m = 20;
7 sd = 10;
8 n = 10000;
9 ylims = [-30 100];
10
11 data1 = (randn(n,1)*sd)+m;
12
13 data2 = [randn(n/2,1)+m-sd; randn(n/2,1)+m+sd];
14
15 v = sd^2;
16 mu = log((m^2)/sqrt(v+m^2));
17 sigma = sqrt(log(v/(m^2)+1));
18 data3 = lognrnd(mu,sigma,n,1);
19
20 data = [data1 data2 data3];
21
22 % bar plots
23 subplot(511)
24 hold on
25 bar(mean(data))
26 errorbar(mean(data),std(data),'.')
27 title(['(\mu)1 = ' num2str(mean(data(:,1))) ', (\sigma)1 = ' num2str(std(data(:,1)))...
28 ', (\mu)2 = ' num2str(mean(data(:,2))) ', (\sigma)2 = ' num2str(std(data(:,2)))...
29 ', (\mu)3 = ' num2str(mean(data(:,3))) ', (\sigma)3 = ' num2str(std(data(:,3)))'])
30
31 % box plots
      
```
- Workspace:** Lists variables in the workspace, including 'ALLCOM', 'ALLEEG', 'ALLERP', 'ALLERPCOM', 'amica\_filename\_input', 'amica\_filename\_output', 'ans', 'baseline\_start\_end\_1', 'baseline\_start\_end\_3', 'channel\_locations\_filename', 'channels\_to\_remove', 'clusters', 'copy\_weights\_interpolate\_avRef\_file...', 'CURRENTERP', 'CURRENTSET', 'CURRENTSTUDY', 'data', 'data1', 'data2', 'data3', 'deleted\_events\_filename', 'design\_name\_icaersp', 'dim', 'dim2', 'do\_clustering', 'do\_multivariate\_data', 'do\_remove\_outside\_head', 'EEG', 'eeglab\_path', 'eeglabUpdater', and 'eeg\_channels'.
- Command Window:** Shows the command 'fs >>' and the output 'New to MATLAB? See resources for [Getting Started](#)'.
- Command History:** Displays the commands entered in the Command Window, including 'load(input\_path\_ERSPs\_outward '\freqs', 'freqs');', 'account for EEG time lag', 'times = times - 100;', 'if ~exist('ALLEEG','var')', 'eeglab;', 'pop\_editoptions('option\_storedisk', 1, 'option\_savetwofiles',...', 'option\_single', 1, 'option\_memmapdata', 0, 'option\_eegobject',...', 'option\_scaleicarms', 1, 'option\_rememberfolder', 1, 'option\_don...', 'option\_checkversion', 1, 'option\_chat', 1);', 'end', 'if isempty(STUDY)', 'STUDY = []; CURRENTSTUDY = 0; ALLEEG = []; EEG=[]; CURRENTSET=[];', '[STUDY ALLEEG] = pop\_loadstudy('filename', study\_2\_filename, 'fi...', 'CURRENTSTUDY = 1; EEG = ALLEEG; CURRENTSET = [1:length(EEG)];', 'eeglab redraw'.

# The MATLAB GUI

- Command Window (execute stuff), Current Folder (find stuff), Editor (write stuff), Command History (old stuff), Workspace (data stuff), Inspector (guess...)
- Dock&Undock, Drag&Drop
- Preferences
  - Color Scheme, Font Size
    - Especially if you code at night.
  - Max column width
    - (Prefs -> Editor/Debugger -> Display **AND** Language)
- "F1" for help, "CTRL-C" to cancel current operation

# MATLAB as a fancy calculator

- Use command window and „Enter“ to execute command
- $+$   $-$   $*$   $/$   $^$
- Parentheses for readability and order of operations
- Colon (:) operator for skipping intervals
- “Clc” to clear the command window

# Variables

- Default: ans
- Storing data in the workspace ("x=1")
- Reassign new values to old variables
- Suppress output using semicolon (;)
- Inspect in the workspace or by typing into the command window
- Whos, clear (x, \*, all)

# Book Recommendation

