INTRODUCTION TO THE MODULE
DAY 1 - INTRODUCTION TO THE COURSE & R BASICS

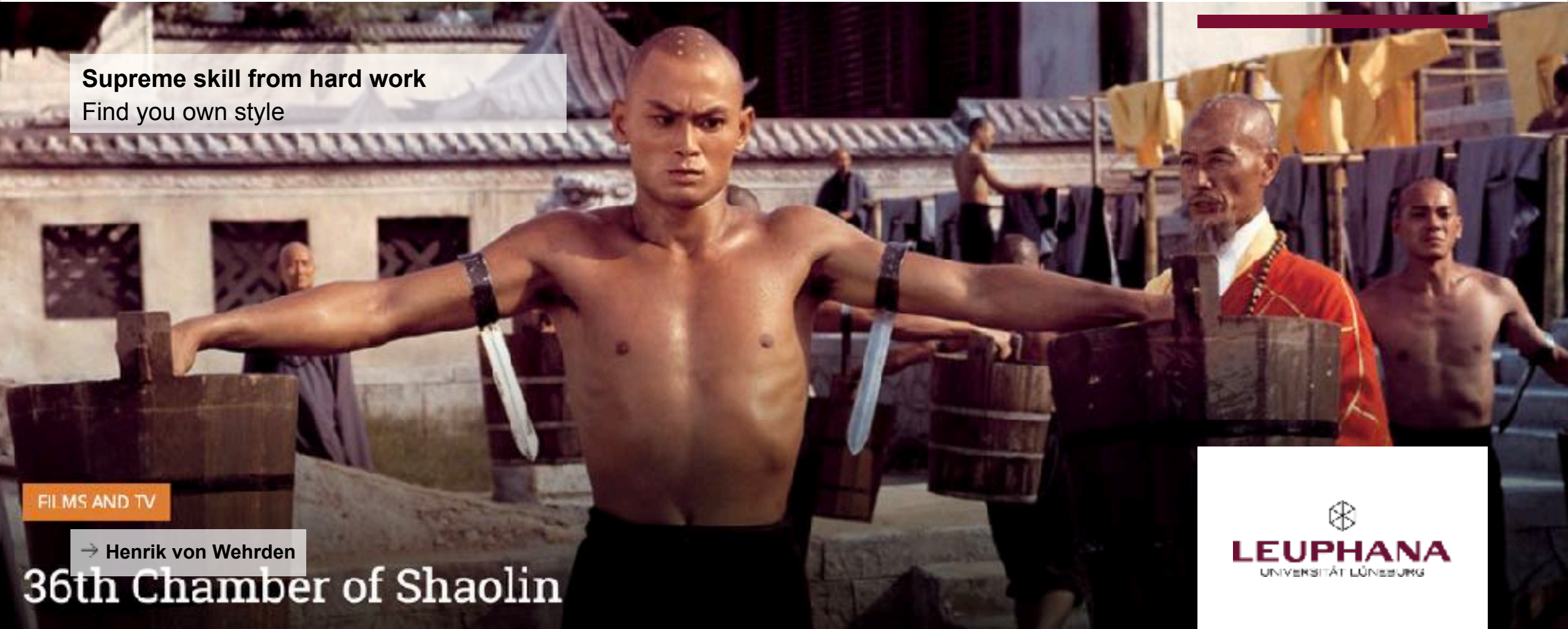
Getting started

Philosophy of the course
Code of conduct
How to participate in this course

Henrik von Wehrden

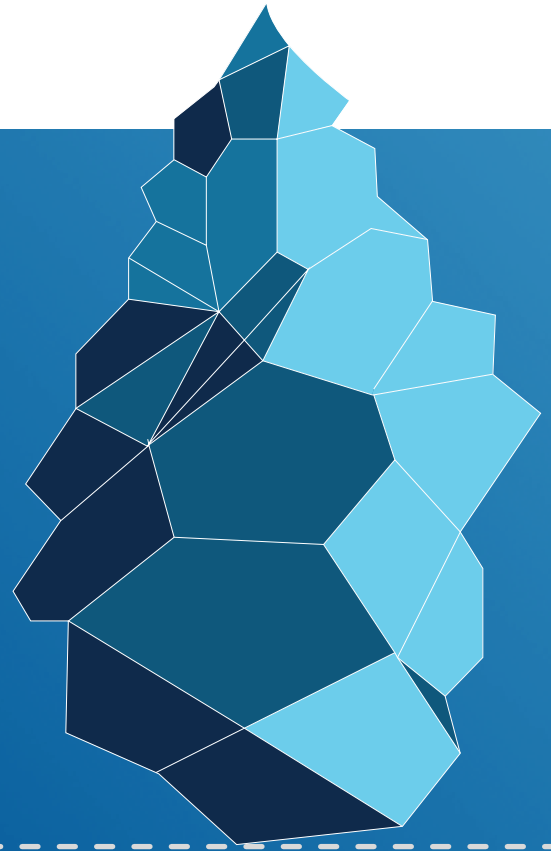
LEUPHANA
UNIVERSITÄT LÜNEBURG

PHILOSOPHY OF THE COURSE
LEARN KUNG FU OF R AND STATISTICS



LEUPHANA
UNIVERSITÄT LÜNEBURG

THE PHILOSOPHY OF THIS COURSE



5%

95%

This course

We compress the equivalent of years of knowledge into this course, providing a structured but surely overwhelming starting point into statistics and R

Statistical experience

In order to gain experience you need to work hard on your own. This course cannot give you experience, yet we believe that learning about advanced statistics you need to spend much time with datasets. This time is about 20 times as much as we have in this course. In order to get prolific, you will need to continue and apply statistics.

WHAT IS ALSO THE PHILOSOPHY OF THIS COURSE

- We am committed to make this course continuously better
- Trust in our experience is essential
- People understand things in their heads
- We provide structure and resources to support this
- I am not aware of any good textbook. All I know are imperfect and disciplinary.
- This course is a stepping stone
- You will not learn about formulas, but are welcome to find this in other ressources

- We can walk with you, we cannot walk for you

WHAT THIS COURSE IS NOT ABOUT

- Your data
- Bayesian analysis or non-linear statistics
- The one and only statistics only you think you need
- You being a master in R and statistics at the end of this course
- Solving your computer problems
- Cherry-picking of the course material

MY CREDO AS A SCIENTISTS

- I am a critical realist
- I struggle with the gap between ontological and epistemological
- I am part of a constructed institution, and seen as a person of power
- I have a post-disciplinary agenda
- I am less serious than people think I am

CODE OF CONDUCT

- Open and respectful communication
- I propose gender neutral language
- We are tolerant against each others flaws
- We are open minded about all branches of science
- We participate actively in this course
- We are empathetic and try to work towards a positive mindset
- We help each other

WHO ARE YOU?

WHO ARE YOU?

- Your degree(s)?
- Main past experience
- One detail about you
- Your knowledge and experience in software for analysing data?

Quick circle of introduction in the course

HOW TO PARTICIPATE IN THIS COURSE

- Active participation
- We all started at some point
- It is ok to not understand everything
- Stick to the timing
- Please ask for clarifications
- A search engine is a good start for definitions
- You can also find formulas in the internet
- Patience with yourself!

STRUCTURAL ELEMENTS OF THIS COURSE

- Short personal input
- Longer input in video, flipped classroom format
- Wiki as an additional repository
- Annotated R examples as a starting point
- Small-group exercises
- Classroom discussions and clarifications

STRUCTURAL ELEMENTS OF THIS COURSE

- Everything combined in Moodle
- Discord for clarifications
- The team and me are there for breakout sessions and questions
- Zoom for exchange and discussions, as well as Discord

THE CURRICULUM OF THIS COURSE

1) R Basics	October 19, 2021	Level 0 The basics
2) Introduction to statistics	October 26, 2021	
3) Data formats and graphics	November 2, 2021	
4) Descriptive stats & simple tests	November 9, 2021	Level 1 Knowledge
5) Distribution and probability	November 16, 2021	
6) Anova	November 23, 2021	
7) Correlations and regressions	November 30, 2021	
6) GLMs	December 7, 2021	Level 2 Experience
7) GLMMs	December 14, 2021	
8) X-Mas slot	December 21, 2021	
9) Model comparision & simplification	January 11, 2022	
10) Ordinations	January 18, 2022	Level 3 Advanced
11) Cluster analysis	January 25, 2022	
12) Advanced analysis	February 1, 2022	

A SHORT WORD ON THE WORKLOAD OF THIS COURSE

Preparation: ~2-4 hours
Lecture: 1.5 hours
Homework: ~2 hour
Seminar: ~1.5 hours
Repeating everything: 2 hours

10 hours/week

140 hours=sufficient time for more practice

GRADING

- Wiki entry on one method (20%)
- 48 hours report in a group (30%)
- Final individual report on a dataset of your choice (50 %)
- Haiku in the final report (5%)

GRADING FOR A WIKI ENTRY

- Not more than 1000 words
- Combining written text with R examples
- Including further links or reading material
- Written to enable people to learn - hence keep it simple

FINAL REPORT (48 HOURS, NOT MORE THAN 10 PAGES)

Narative of the report		10
Visualization, Plots and their supporting role		20
Analysis Steps	Apply models, conduct tests, etc.	30
Structure of the report	Sections and structures are coherrent	10
Methodological Limitations	Disclaimer on the value of the analysis	10
Writing Style	Sentences are understandable	10
Markdown appendix	Report 5 pages, submitted in Markdown	10
Haiku	Can be in English	5

INDIVIDUAL REPORT (NOT MORE THAN 5 PAGES)

Narative of the report		10
Visualization, Plots and their supporting role		20
Analysis Steps	Apply models, conduct tests, etc.	30
Structure of the report	Sections and structures are coherrent	10
Methodological Limitations	Disclaimer on the value of the analysis	10
Writing Style	Sentences are understandable	10
Markdown appendix	Report 5 pages, submitted in Markdown	10
Haiku	Can be in English	5

R SURFACE

- https://www.tidyverse.org

- We will build on tidyr, dplyr and ggplot2
- If you want to use other things from the tidyverse, that is totally fine. R is a language, and languages have dialects.

- Discord! Slack?
- https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

HOMEWORK

- Coordinate yourself in the group
- Start brainstorming on your learning goals
- Finalise the code of conduct
- Start with https://swirlstats.com

SUMMARY

- You will get in introductionary glimpse at statistics in this course
- We are committed to a respectful learning environment
- We acknowledge diversity and different experiences
- We are committed to participate actively in this course

R SURFACE

```
Read data table
read.table
read.csv

set working directory
setwd
```


R SURFACE

```
nums = 1:10

nums + 1
```

```
 [1]  2  3  4  5  6  7  8  91011
```


R SURFACE

```
matrix
```

```
rmat = matrix(rnorm(15),5,3,
  dimnames=list(NULL,c('A','B','C')))
```

```
colnames
rownames
```


R SURFACE

```
1:10

seq(1,10)

seq(10,100,5)

seq(10,by=5,length=10)
```


R SURFACE

```
fert = c(10,20,20,50,10,20,10,50,20)
   fert =factor(fert,levels=c(10,20,50),
  ordered=TRUE)

fert
   [1] 10 20 20 50 10 20 10 50 20
   Levels: 10 < 20 < 50
```


R SURFACE

```
lets = sample
(letters,size=100,replace=TRUE)

lets = factor(lets)

table(lets[1:5])
```


R SURFACE

```
nums = c(12,9,8,14,7,16,3,2,9)

nums > 10

nums > 10 & nums < 16
which(nums>10)

nums[nums > 10] = 0
```


R SURFACE

```
method1 = c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)

method2 = c(1,2,2,3,2,2,1,3,3,3,2,4,1,4,4,3)

tt = table(method1,method2)

tt
```


R SURFACE

```
iris
str(iris)
names(iris)
head(iris)
summary(iris)

mean(iris$Petal.Width)
table(iris$Petal.Width,iris$Species)
```


R SURFACE

- https://www.tidyverse.org

- We will build on tidyr, dplyr and ggplot2
- If you want to use other things from the tidyverse, that is totally fine. R is a language, and languages have dialects.

- Discord!
- https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

THE MODE OF CODING

Your computer

Your data

Your scripts

Your sections

Your lines

Your words

YOUR COMPUTER

Hardware: Split screen, keyboard, touchpad,

The specs do not matter now

Language setting: Mac vs. PC, browser, number settings etc.

Software suite: R Studio, Packages, Notion, Discord, Slack, No social media, Text writing, etc.

Harddrive: The desktop is no good place for data. Get a long term structure. Clean up.

Backups: At least three backup plans.

Get a second brain

YOUR SCRIPTS

Clearly naming codex, including date

Save the scripts with the data

Get a beginning section that explains the current steps

One script per project. More than 10 lines, less than 1000

Get section instructions and summaries of the steps you made

Ask other for feedback. Learners are teachers, teachers are learners.

CODE SECTIONS

One section per analysis or data deriving step

Sections are like poems

Hashtag lines above sections

Hierarchical nestes structure

Hashtags after each line

Tab stopped structure

Avoid line breaks if possible

Do not be impressive

TAKEN TOGETHER

- There are basic repositories that you need to dive into to learn the basics of R and coding
- Getting your work environment, your computer and your coding in order are first steps
- You need to practice daily, exchange continuously, and work towards at least 1000 hours of coding to become a master
- If you are committed, your knowledge may grow into experience
- Out of experience, wisdom may arise
- Remember that your have a privilege and a responsibility

HOMEWORK

- Coordinate yourself and try to form groups
- Start brainstorming on your learning goals
- Reflect on the code of conduct
- Start with https://swirlstats.com

CONTACT

Henrik von Wehrden | Faculty of Sustainability
Universitätsallee 1 | 21335 Lüneburg
Fon 04131.677-1671 | henrik.von_wehrden@leuphana.de