INTRODUCTION TO
DATA ANALYSIS

MAIN COURSE MATERIAL

▸ **course website** (link also on StudIP)

  ▸ https://michael-franke.github.io/IDA-2019/

  ▸ slides, homework etc. will appear here

▸ **course notes as web-book** (link also on StudIP)

  ▸ https://michael-franke.github.io/intro-data-analysis/

  ▸ main reading
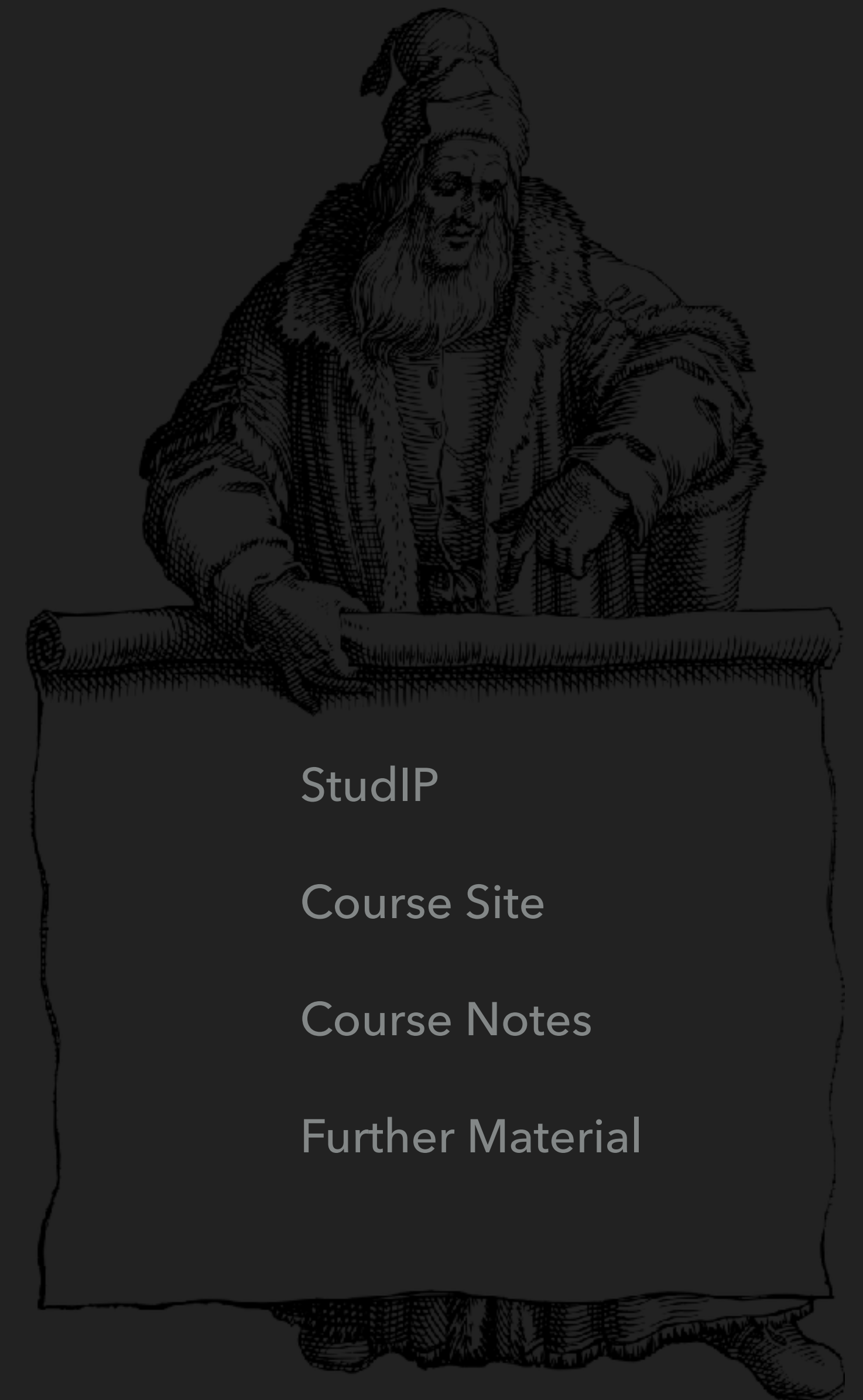
StudIP
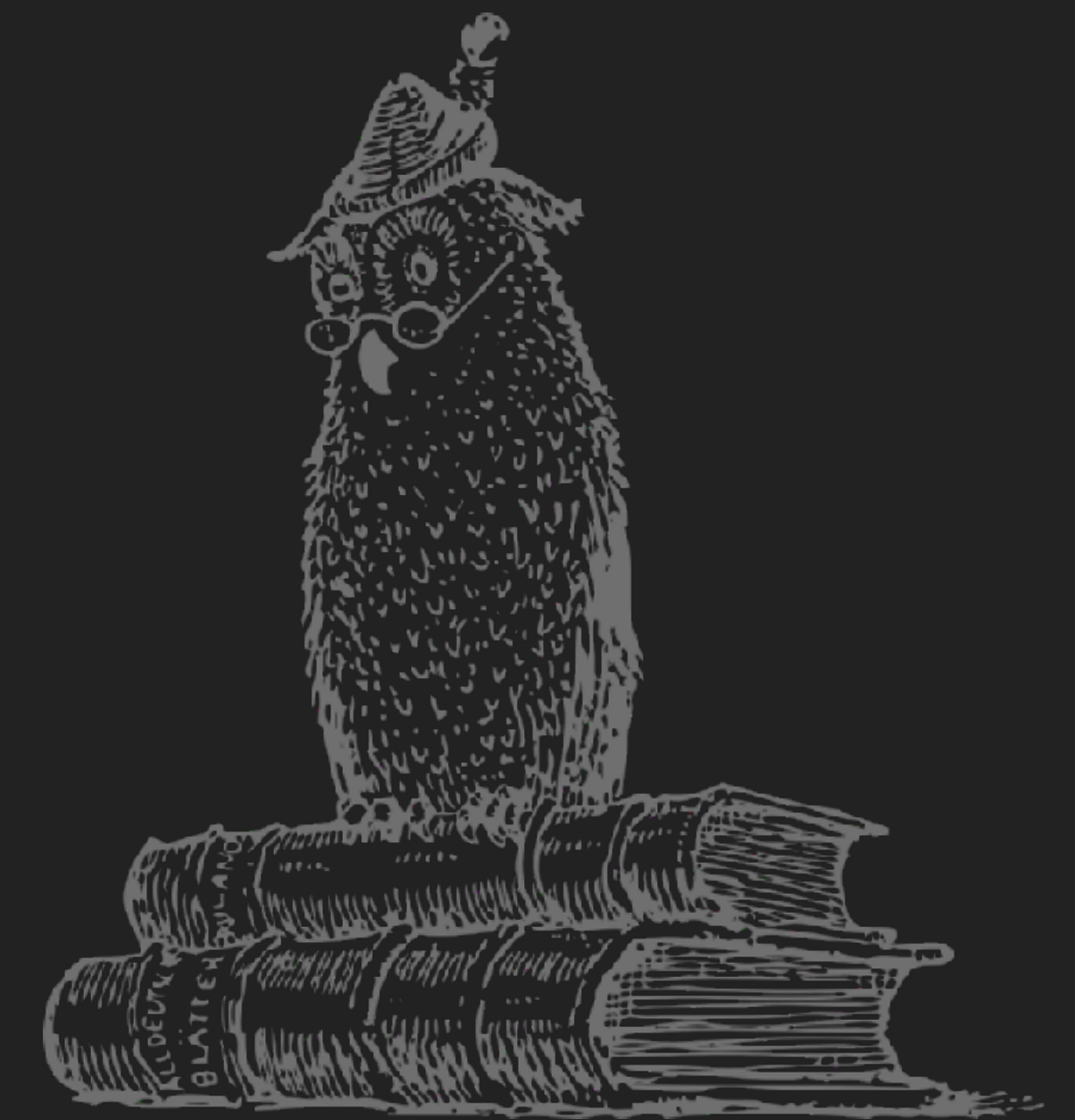
Course Site

Course Notes

Further Material

LECTURES

▸ make sure to catch both lectures each week

  ▸ Wednesday, 10:15-11:45 (66/E33)

  ▸ Friday, 12:15-13:45 (32/102)

▸ prepare reading in advance (see schedule on website)

▸ reread chapter after lecture

TUTORIALS

▸ tutorials give extra background and practical exercises (tutors vary w/o notice)

▸ everybody should catch at least one tutorial per week

▸ tutorial times and locations (also on StudIP)

    ▾ Monday, 10:15-11:45 (66/E34)

    ▾ Tuesday, 8:15-9:45 (66/E33)
    [this tutorial will not take place on Nov 5]

    ▾ Tuesday, 12:15-13:45 (32/107)

IDA-2019 TEAM

▸ tutors

  ▸ Tallulah Jansen, Nina Mainusch, Maria Pershina,
    Jona Carmon, Taher Habib, Marc Viladrich

▸ web-book

  ▸ Florence Bockting, Tobias Anton

▸ additional support

  ▸ Noa Kallioinen, Ann-Christin Meisener

COURSE REQUIREMENTS

▸ required for passing:

1. passing grade on homework assignments

   ▸ weekly assignments to be submitted in groups of three

2. passing grade on final exam

   ▸ 4h written, "open-book" (bring hand-written notes)

▸ see course website for more information
https://michael-franke.github.io/IDA-2019/grading/

HOMEWORK

▸ HW issued on Friday evening (after lecture)

▸ HW due Friday next week at noon (before lecture)

▸ submit electronically via StudIP

  ▸ upload to your group's folder

  ▸ as Rmd or PDF (LaTeX, hand-written scan)

▸ no detailed comments as corrections but sample solution

LEARNING GOALS

▸ ability to explore data sets in hypothesis-driven manner

▸ manipulate & visualize data

▸ understand logic of statistical inference (frequentist vs Bayesian statistics)

▸ ability & confidence to critically assess DAs in research papers

▸ ability & confidence to tackle your own DA for an experimental BSc thesis

TECHNICAL SKILLS YOU WILL ACQUIRE IN THIS COURSE

▸ basics of R & `tidyverse`

   ▸ write your own DA scripts
   (manipulation, visualization, statistical analyses)

▸ reproducible writing in Rmarkdown

▸ glimpse at probabilistic programming languages
(WebPPL, greta)

▸ first contact with generalized linear models

WHAT YOU SHOULD NOT EXPECT OF THIS COURSE

▸ details of common algorithm for statistical computation (MCMC, optimization, …)

▸ emphasis on history and/or philosophy of statistics

▸ reactive practical competence in statistics

  ▸ "Our aim is understanding."

  ▸ "We do not teach tricks!"
     "We do not share recipes!"

COMPARISON OF RELATED COURSES

▸ **statistics @ Psych**

  ▸ focus on frequentist methods

  ▸ SPPS instead of R

▸ **probability theory @ Math**

  ▸ focus on theory not applications

  ▸ math not computer science

  ▸ foundational issues

▸ **neuroinformatics @ CogSci**

  ▸ more focus on theory & math

  ▸ applications in neuroscience / ML

▸ **intro data analysis @ CogSci**

  ▸ focus on computation

  ▸ applications in behavioral psych

  ▸ frequentist & Bayes

SCHEDULE (PRELIMINARY)

| week	Content of tutorials	Wednesday lecture	Friday lecture	HW issued
44			**Course overview** Chapter 1	
45	installing R & packages (Stan, tensorflow)	**Intro to R** Chapter 2	**Data & data handling** Chapter 3	HW1
46	using R, data handling / wrangling	**Data wrangling** Chapter 4	**Summary statistics** Chapter 5	HW2
47	more R, wrangling, summary stats	NO LECTURE	**Data plotting** Chapter 6	HW3
48	plotting, more R exercises	**Probability basics** Chapter 7	**Frequentism vs Bayes** Chapter 8	HW4
49	probability calculus, Bayes rule	**Statistical models** Chapter 9	**Parameter Inference 1** Chapter 10	HW5
50	simulations on error control, calculations with Bayes rule	NO LECTURE	**Parameter Inference 2** Chapter 10	HW6
51	sampling-based approaches using R, WebPPL, greta	**Classical testing 1** Chapter 11	**Classical testing 2** Chapter 11	HW7

week	Content of tutorials	Wednesday lecture	Friday lecture	HW issued
2	binomial test, t-test, (maybe ANOVA), …	**Classical testing 3** Chapter 11	**Model comparison** Chapter 12	HW8
3	model comparison, Bayes factors, LR test, AIC	**Bayesian hypothesis testing** Chapter 13	**Model criticism** Chapter 14	HW9
4		**Simple linear regression**	**Generalized regression**	HW10
5		**Generalized regression**	**Hierarchical regression**	HW11
6		Q&A	**Final exam**	
7		**Cognitive models in data analysis**	**TBA**	

check course website for updated schedule

FLAVORS OF MODERN STATISTICAL ANALYSIS

FREQUENTISM

objective probability :: tests :: *p*-values :: error control
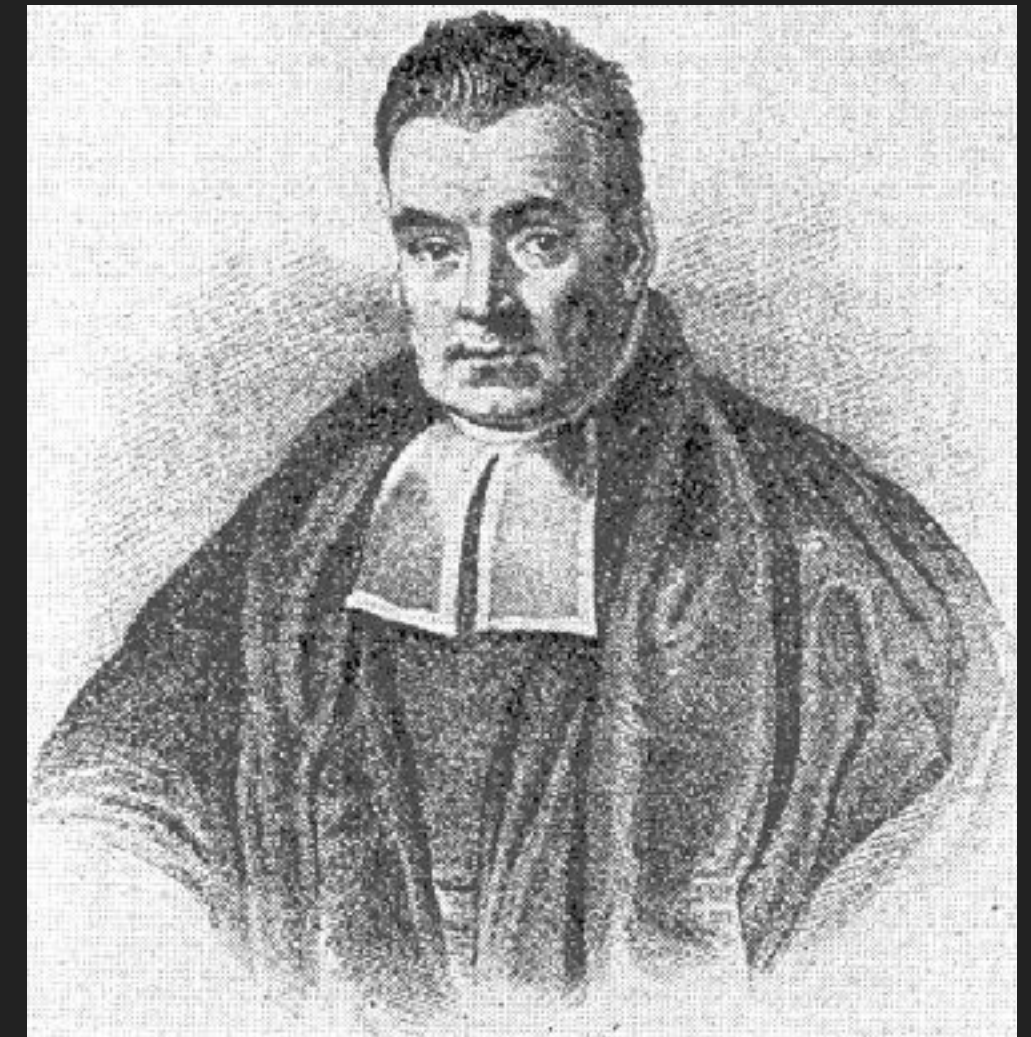
BAYESIANISM

subjective probability :: beliefs :: updates
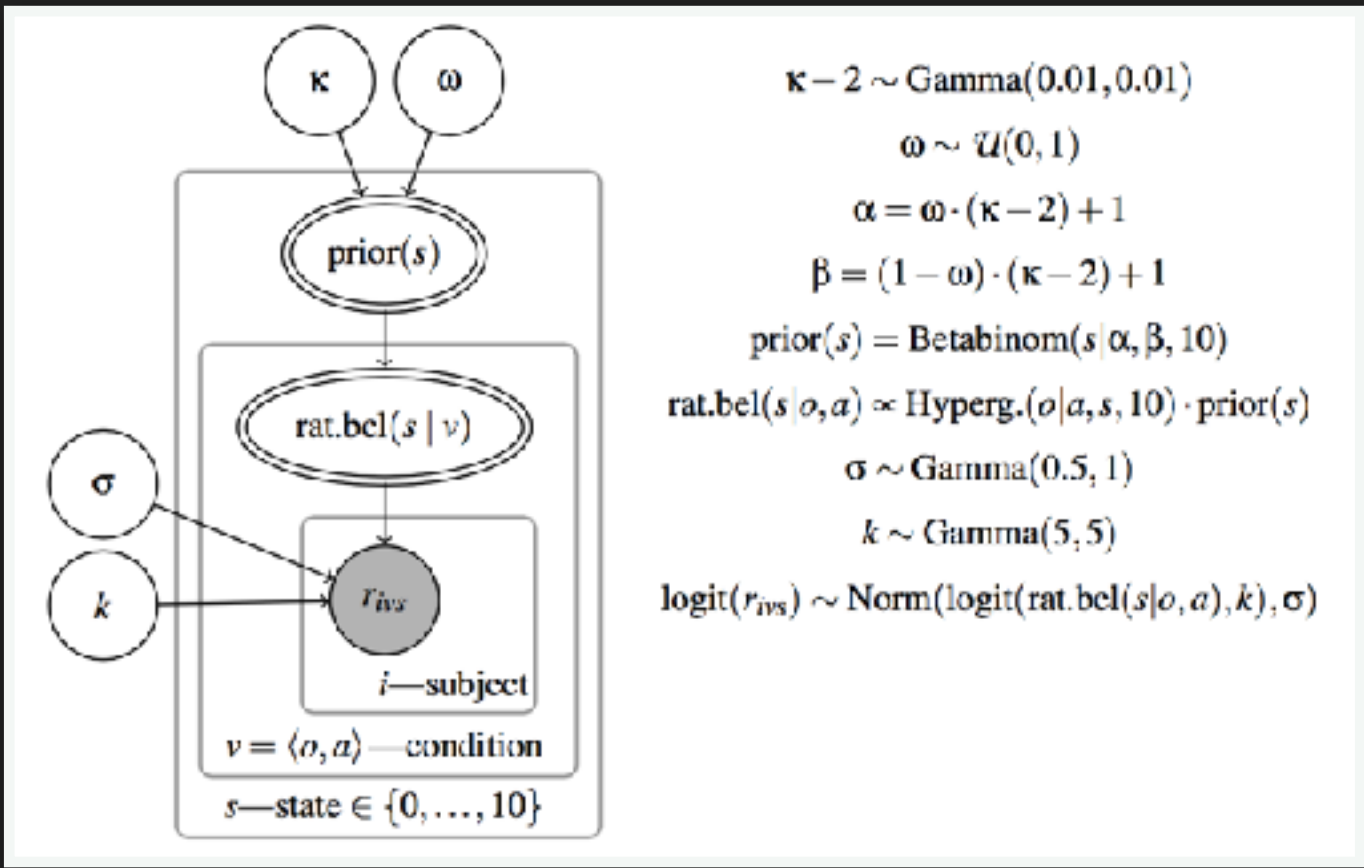
STATISTICS AS POTTERY

STATISTICAL INFERENCE

adresses

RESEARCH QUESTION

?

yields

STATISTICAL MODEL

EMPIRICAL DATA

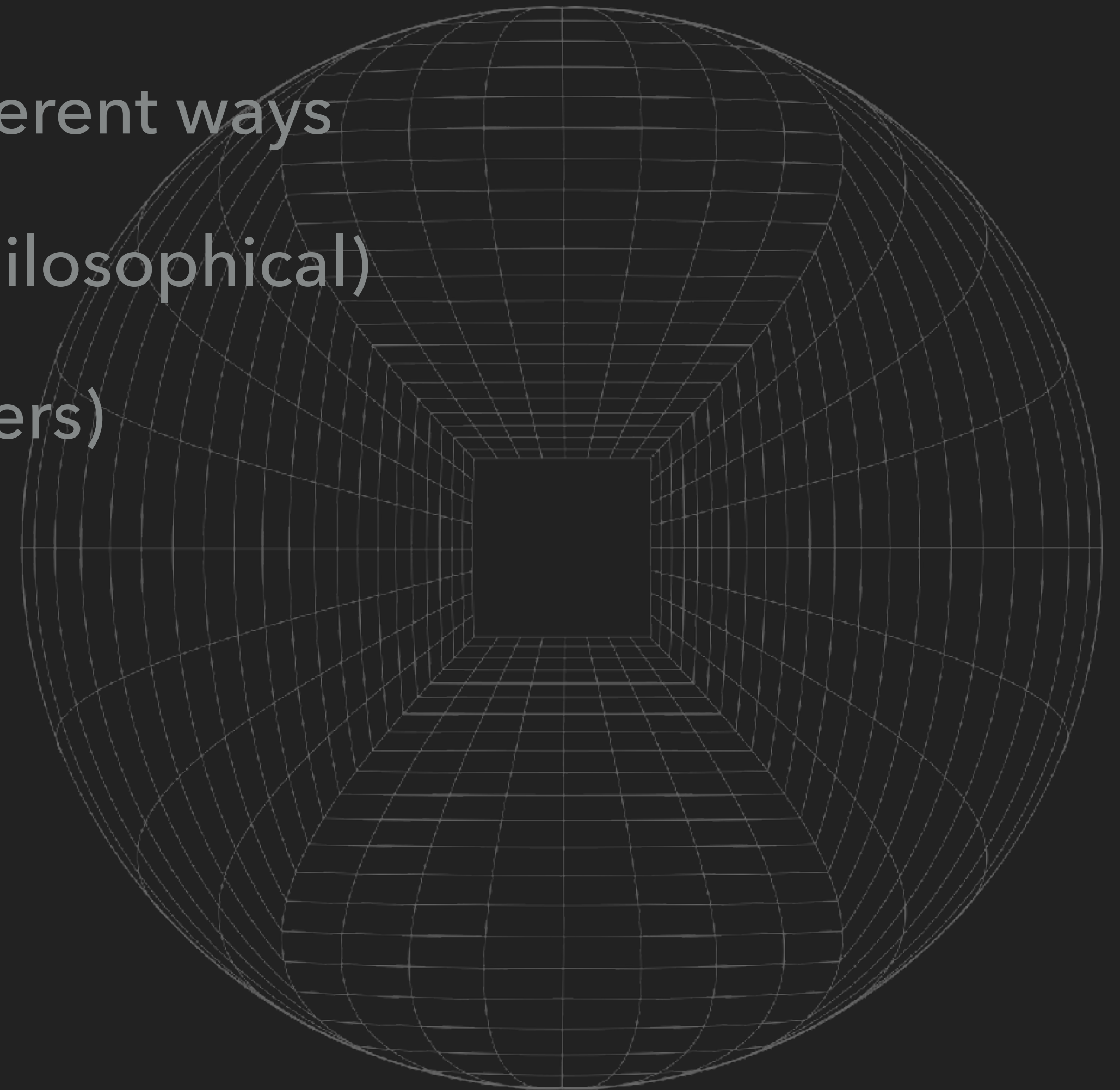motivates

fuels

UNDERSTANDING THROUGH COMPUTATION

▸ statistical concepts can be understood in different ways

  ▸ in terms of their motivation (practical or philosophical)

  ▸ as a mathematical concept (elegance matters)

  ▸ through implementation as algorithms

  ▸ by exploring computational simulations

SELF-ANALYSIS

▸ aspiring analysts should undergo analysis themselves

▸ running examples based on online experiments

    ▸ whoever wants can participate

    ▸ collected data will be used in exercises and HW
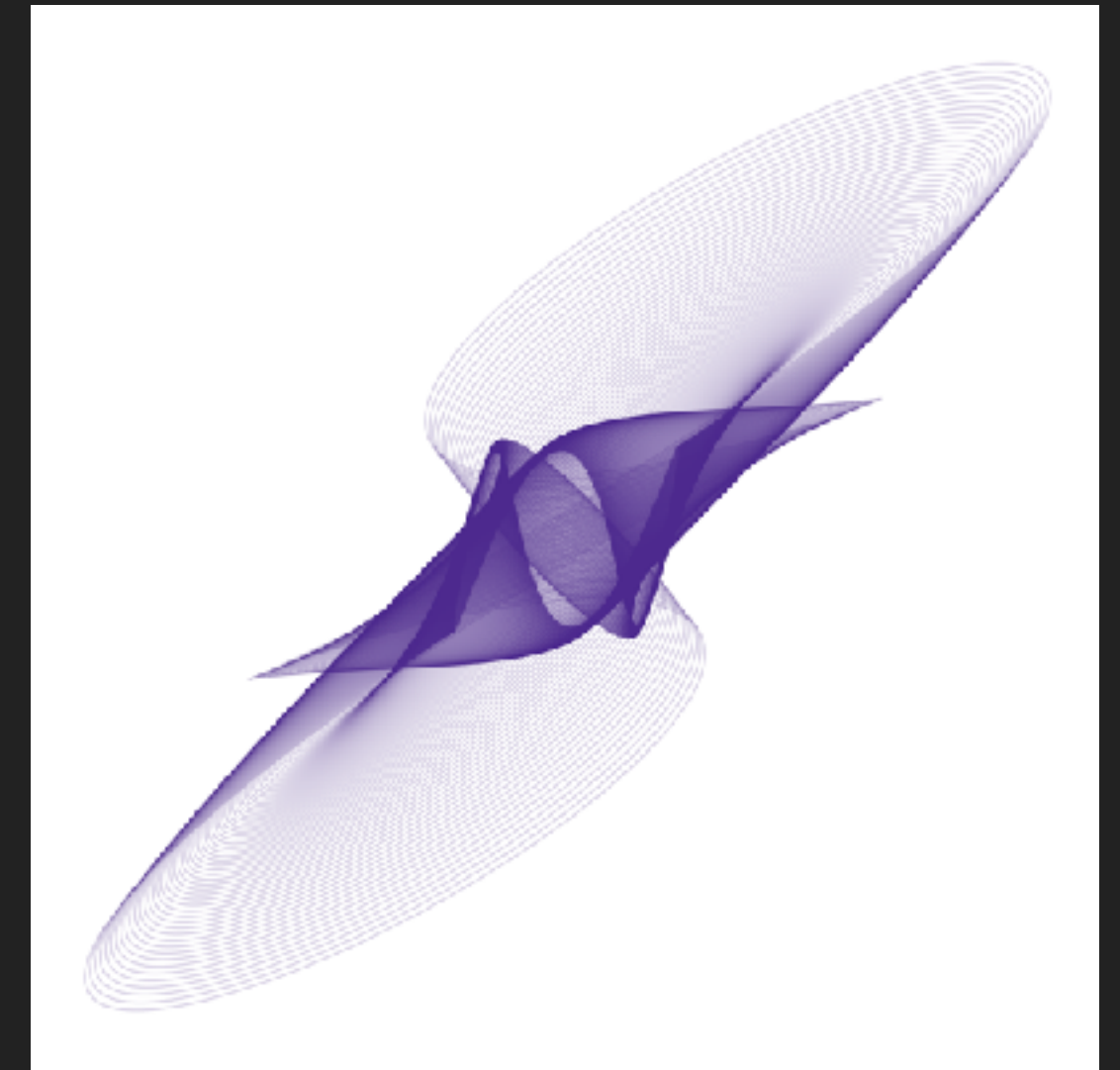
FURTHER STUDY MATERIAL

▸ appendix chapter A of course material

▸ top pick on stats books:

  ▸ Ben Lambert (2018) "A Student's Guide to Bayesian Statistics"

  ▸ Bodo Winter (2019) "Statistics for Linguists: An introduction using R"

CONTACT

please direct all communication to

 Tallulah Jansen <taljansen@uni-osnabrueck.de>

HOMEWORK FOR NEXT CLASS

▸ read Chapter 1 of course notes

▸ install all necessary software as described in Chapter 1.5

▸ visit a tutorial next week to get help with installation

   ▸ no tutorial on Tue at 8:15-9:45 on Nov 5

▸ prepare Chapter 2