

Machine Learning and the Social Sciences

04/06/2017

Prof. Dr. Jörn Grahl

grahl@wiso.uni-koeln.de



About me

- **Faculty of Management, Economics and Social Sciences**
- **Information systems**
- **Data Analytics, Digital Transformation**



This talk

- Relationship between ML and Soc. Sci.
- Hypothesis:
 - ML and Soc Sci should work together, but don't.



Who, what?

- **Social sciences = „Wiso-Faculty“**
 - Marketing
 - Economics
 - Psychology
 - Social Scientists
 - ...
- **ML = ML + data-driven research (Data Science, computer science)**



Agenda

- **Warning**
- **Top 5 reasons for not working together**
- **Top 3 reasons for collaboration**

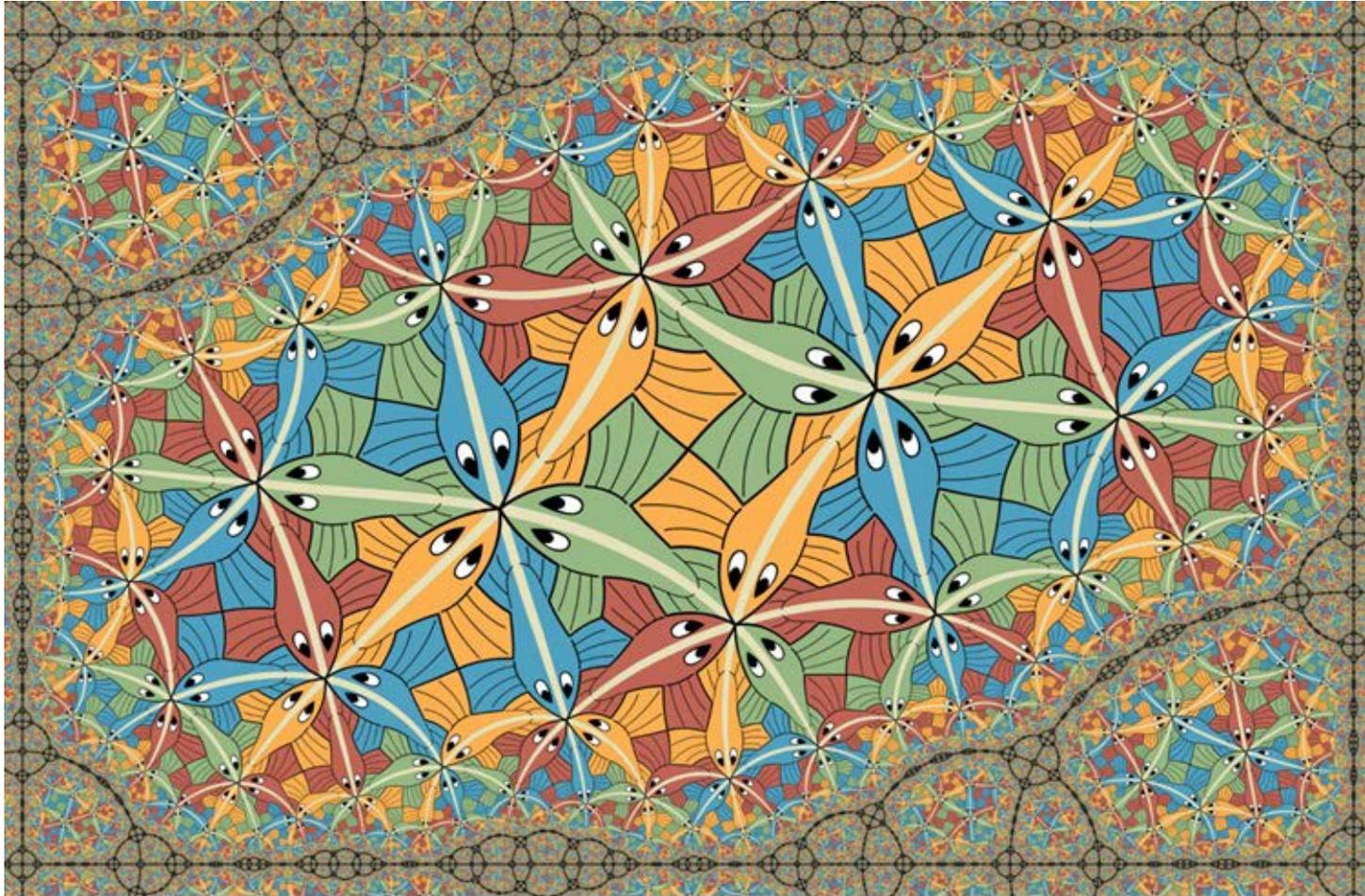


Agenda

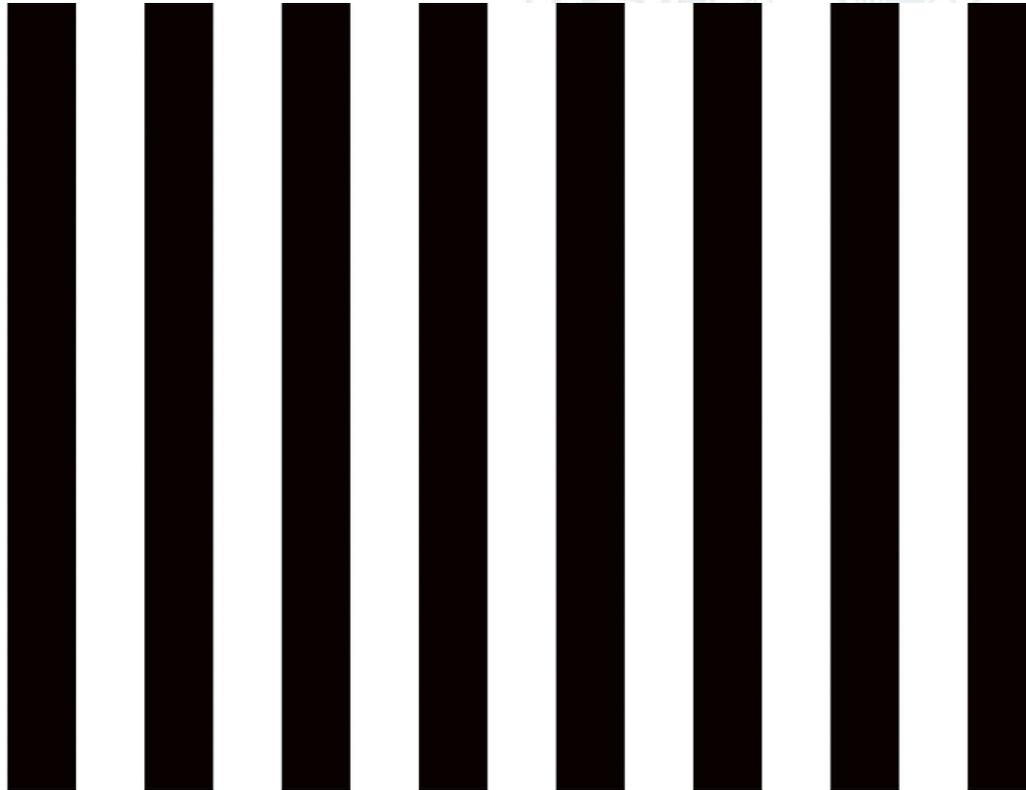
- **Warning**
- Top 5 reasons for not working together
- Top 3 reasons for collaboration



Maybe, ML in the Soc' Sci looks like this

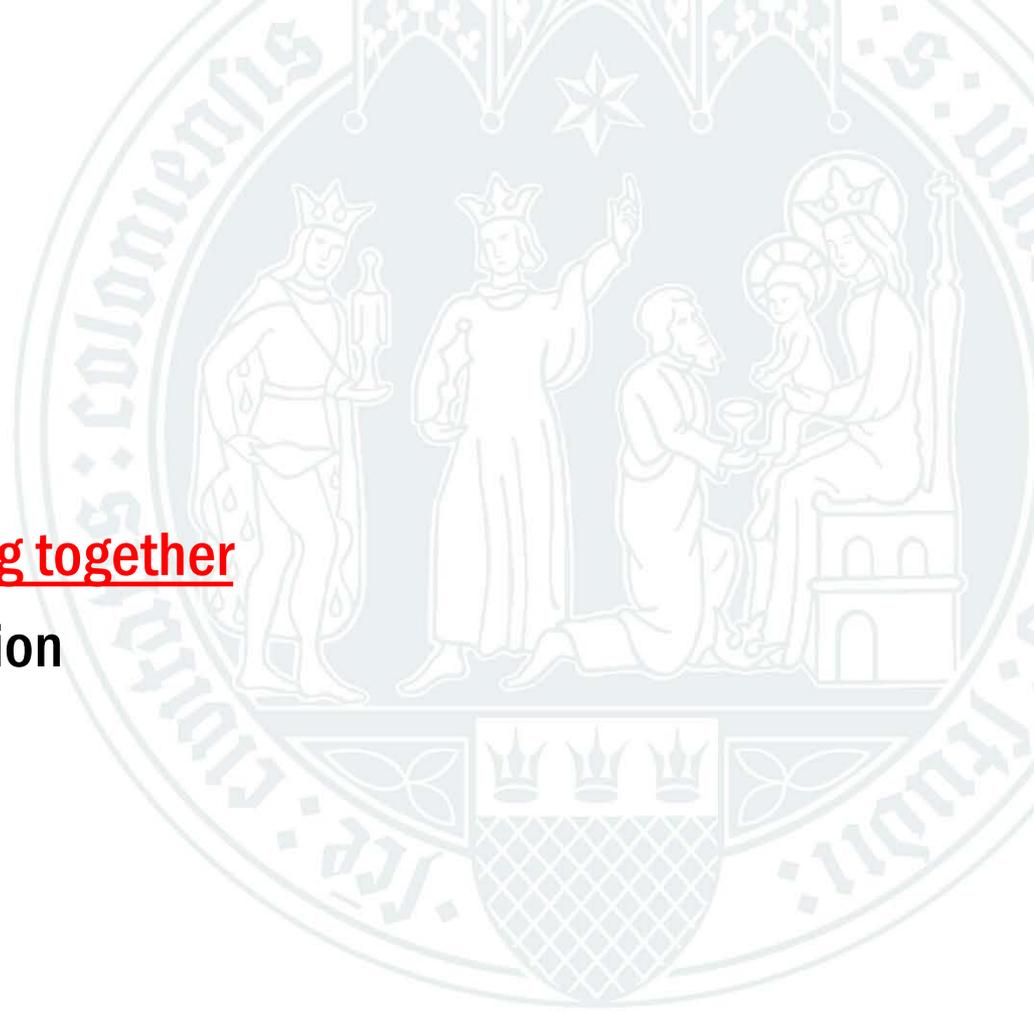


I will simplify a bit



Agenda

- Warning
- Top 5 reasons for not working together
- Top 3 reasons for collaboration



1: Observable means boring

- ML ist applied to large datasets about things that are usually observable
- Otherwise: no large dataset, no ML
- Most observable data is (unfortunately) uninteresting
- Deep data comes in small samples: ML not useful



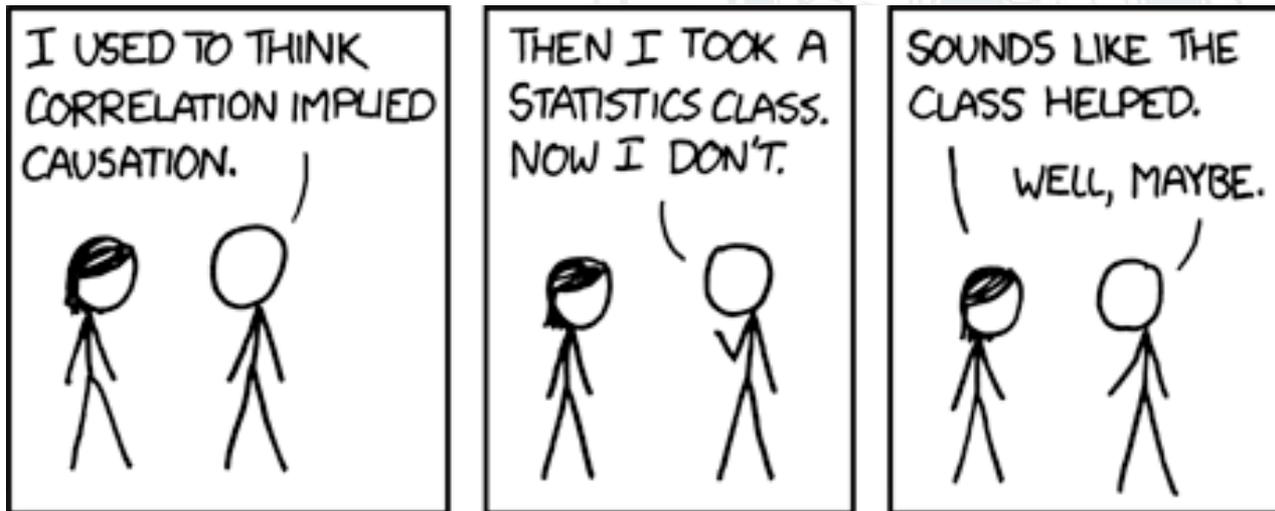
„Looking good“

2: Control the data-generating process

- How data is generated is much more important than size
- Goal: Causal inference, causality, causal effects
- Will allocate time and money and energy to get a counterfactual
- Experiments, in lab, in the field, on digital platforms
- ... and generate their own data;
 - small n
 - tailored towards specific statistical tests, not ML



ML does not solve statistical challenges



3: Size is not everything

- Having more rows is nice (increases statistical power)
- Having more columns is nice, (estimate heterogenous effects, interactions)
- But with limited research resources:
 - If the choice is between creating a counterfactual or more data, the counterfactual will mostly win



4: Prediction can mean many things

- ML is great for prediction
- ML predicts from correlations
- This is not always attractive for a social scientists
- Want to predict on the basis of theory that explains causal relations



5: Typical research questions are different

- **Social Science**
 - How does A influence B, everything else being equal
- **Machine learning**
 - How does A „change (with)“ B, everything else moves freely



Summary

The old-school

- school of thought
- grad school course system
- researcher
- mindset

is happy without ML



This is slowly changing

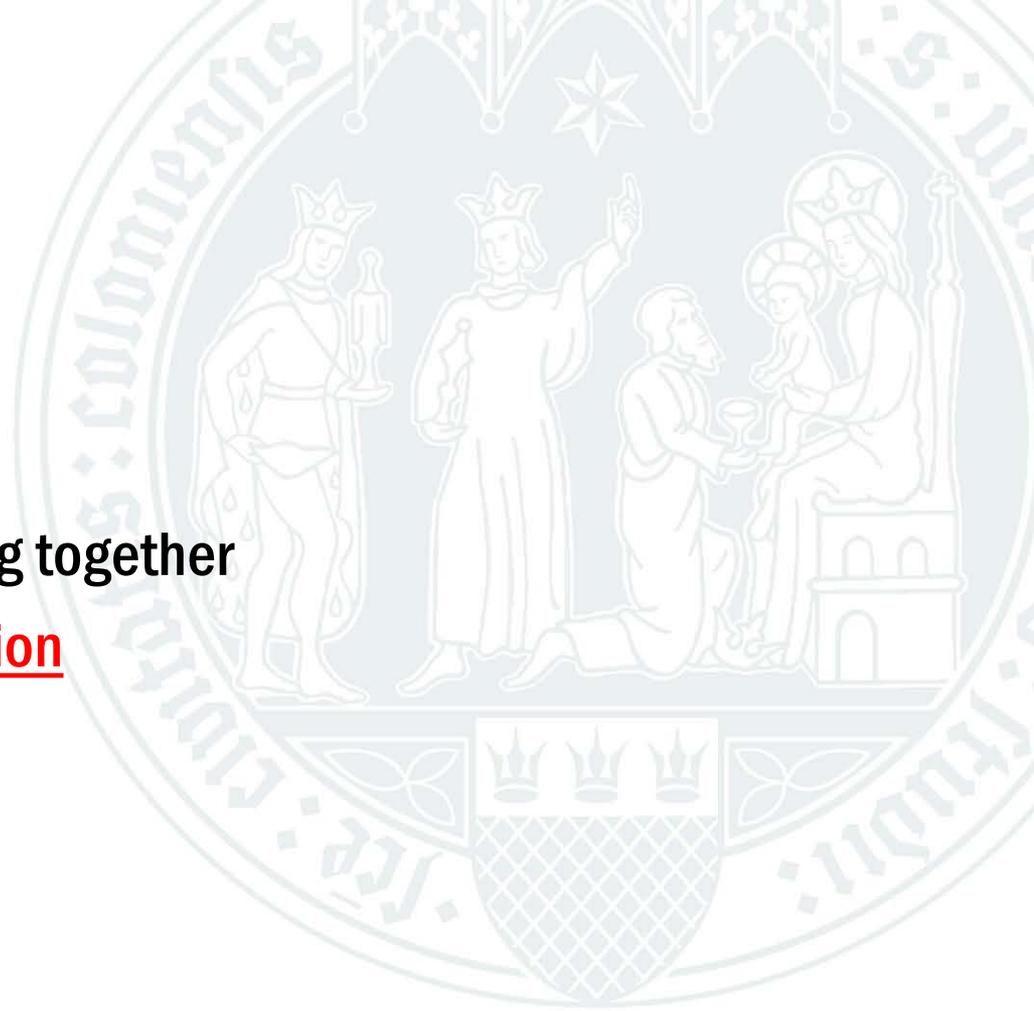
Working together with data people

- ... is a competitive advantage today
- ... is essential in the future



Agenda

- My perspective
- Top 5 reasons for not working together
- Top 3 reasons for collaboration



1: Fewer constraints on topic choice

- Dealing with data is no longer a limiting factor
- You can choose very different research projects
- Example: large randomized controlled field experiments on digital platforms



2: CS and ML can help with causal inference

- Natural experiments can be hidden in a large dataset
- You could use ML to predict the counterfactual



3: Fast model building and testing

- **Very quickly:**
 - Find complex correlations
 - Outside the constraints of theory
 - generate hypotheses “automatically”



Note: Computational social science

- Quant and computational methods for large scale social data
- How to use large datasets with computers is currently hot
- ML is next step, maybe.



Conclusion

- **(Grad)-students: learn Econometrics and ML**
- **Data Science (which includes ML) should become a university-wide (faculty-spanning) priority**

