

**Get off my lawn!**

**How to avoid trespassing in  
multidisciplinary collaboration.**

*Erik-Jan van Kesteren*

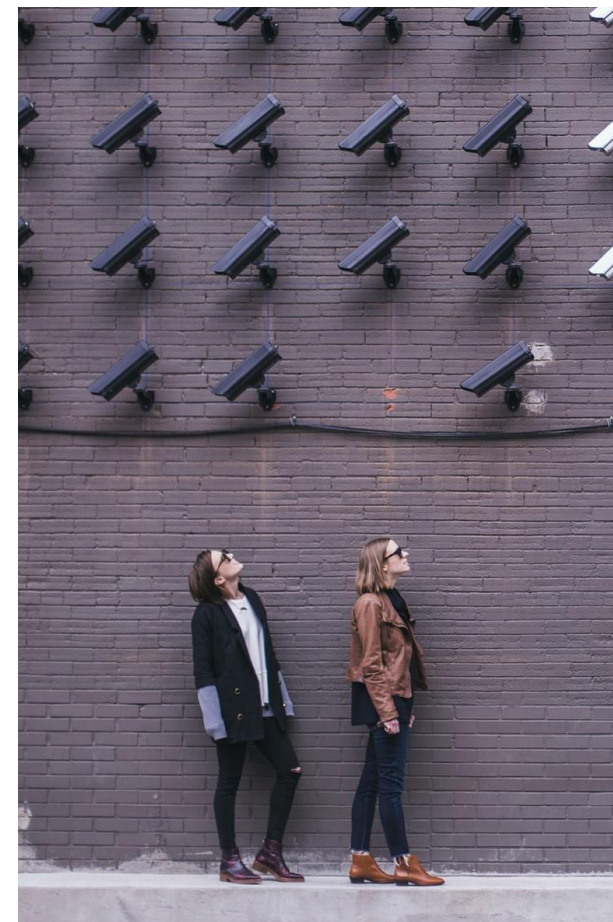


# Today

**Three  
statistics  
examples**

**What went  
right or  
wrong?**

**Guidelines  
to avoid  
trespassing**



stock images in this presentation from [unsplash.com](https://unsplash.com)

# About me

## Education

- BA Social Sciences
- MSc Organisational & Social Psychology
- MSc Methodology & Statistics
- PhD Statistics / Data Science

## Work

- Universitair Docent (Assistant Professor)  
Human Data Science at Utrecht University
- Team lead of *Social Data Science team*  
(i.e. a lot of trespassing)
- Ex-board member of YS!!



# Topics I worked with

## **Regression**

Structural equation modelling

Mediation analysis

Regularization

High-dimensional data

## **High-performance computing**

Data visualisation

Computation graphs / optimization (tensorflow / torch)

## **Privacy / Federated learning**

Software development

## **Bayesian workflow**

Geostatistics / geodata

## **Data science pipelines / open science**

...

# Example 1: Formula One Analysis

## Sports statistics



# Formula One Analysis

## Question

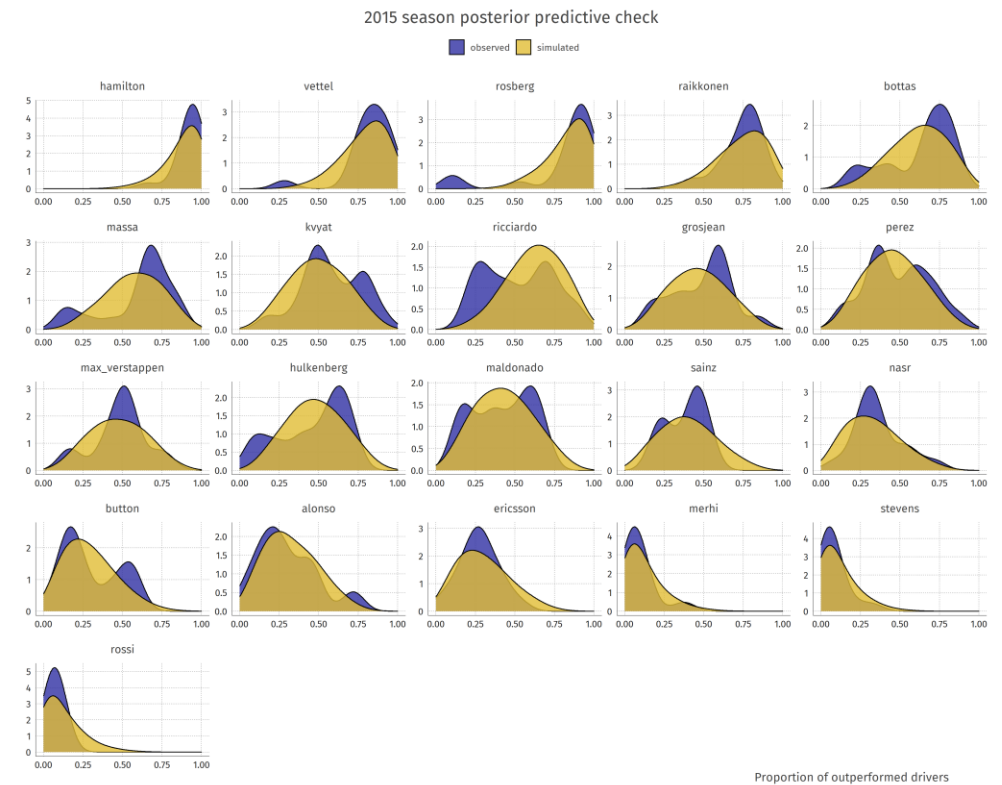
- Who is the best Formula One driver (skill, talent)?

## Problem

- Success is in large part determined by the car
- We want driver skill conditional on car power

## Solution

- Bayesian multilevel Beta regression model
- Outcome: proportion of competitors defeated in the race
- Log odds-ratio per driver, 0 means you are average

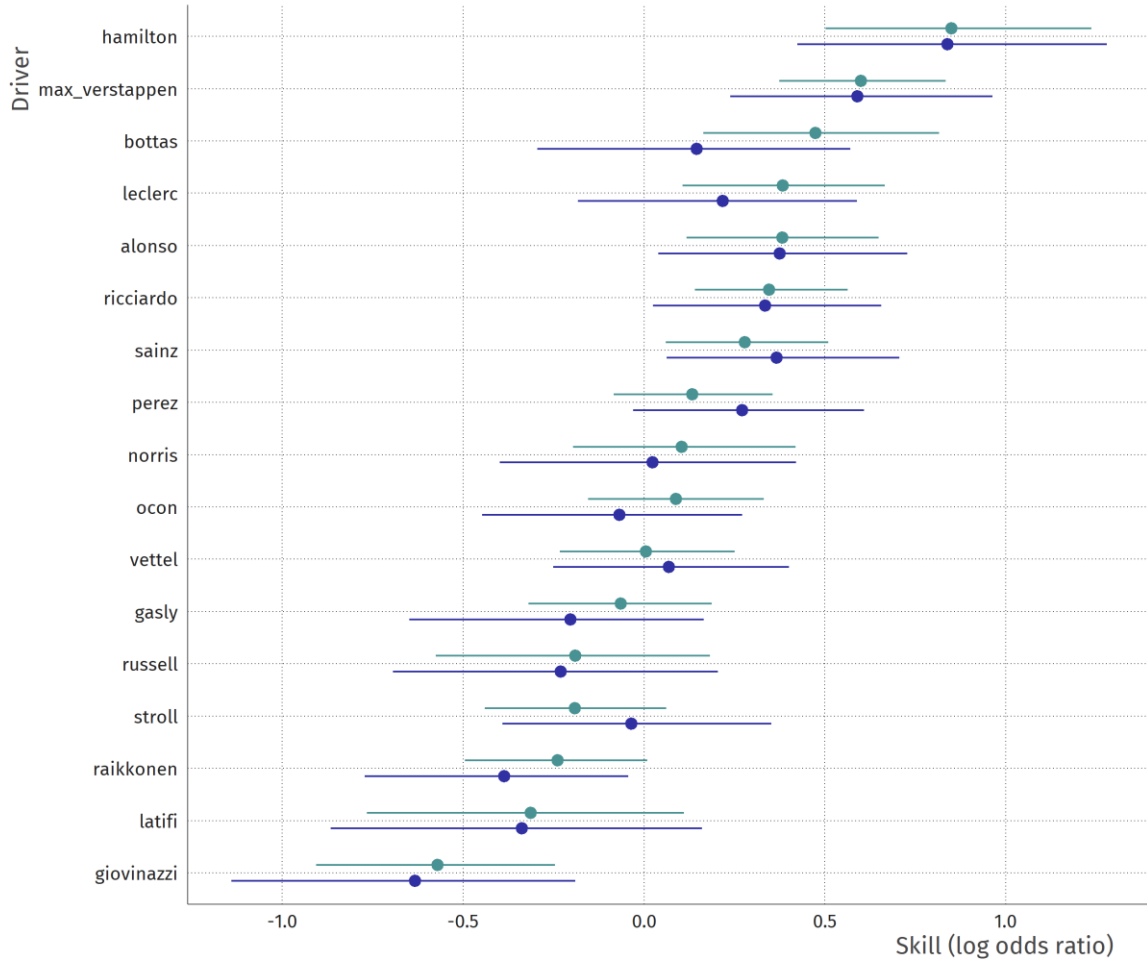




### F1 driver skill

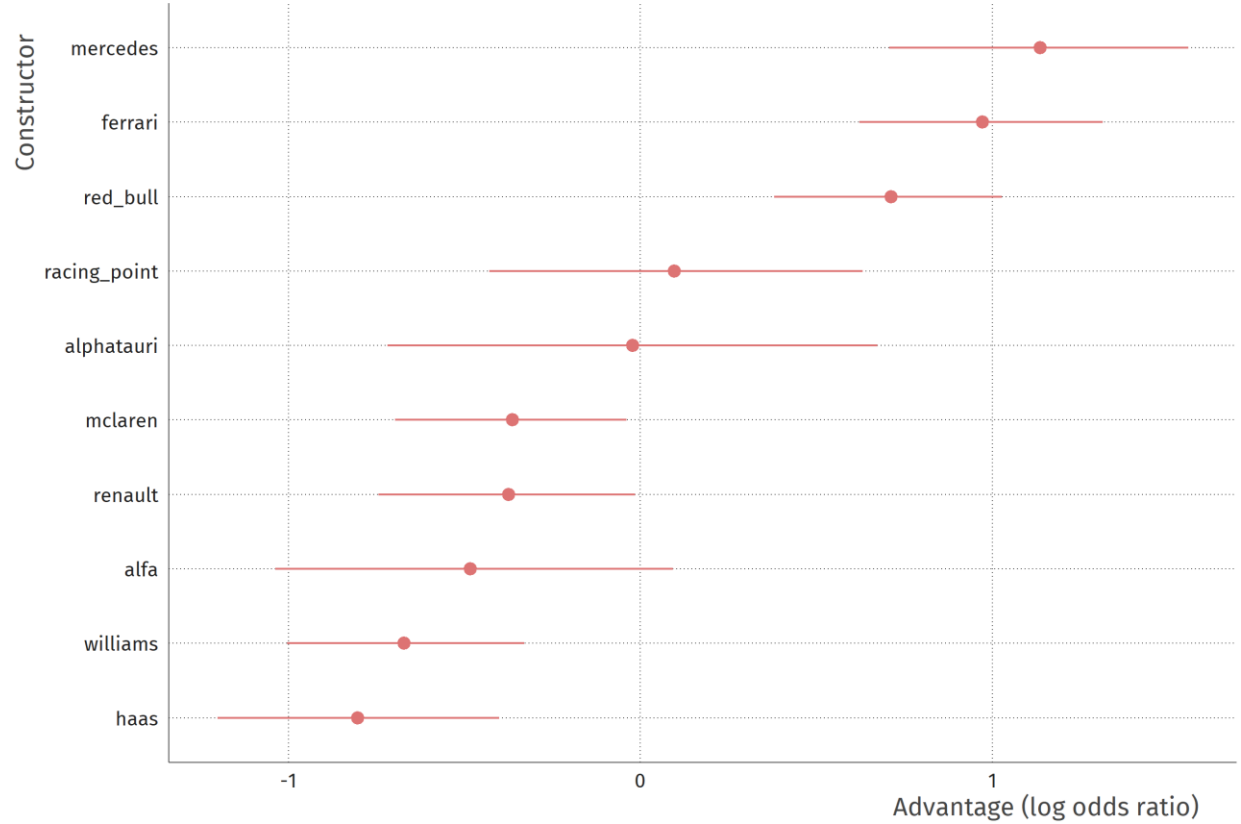
Average hybrid-era (2014-2020) driver skill, accounting for yearly constructor advantage.

● Dry race ● Wet race



### F1 constructor advantage

Average hybrid-era (2014-2020) constructor advantage, accounting for driver skill & constructor form.



# What went well?

## I found a collaborator!

Tom Bergkamp, study friend & PhD in sports analytics / talent identification

## Tom explained the rules

- Reorganised paper (e.g., more focus on question rather than method)
- Identified what is new/valuable to the field
- Rewrote to correct language (e.g., talent -> skill)



The screenshot shows the profile page for Tom Bergkamp on the University of Groningen website. The header is red with the university logo and name. The navigation bar includes Home, Profiles (highlighted), Research Units, Research output, Projects, Datasets, Prizes, and a menu icon. The profile section features a photo of Tom Bergkamp, his name, and his credentials: MSc, PhD student, Faculty of Behavioural and Social Sciences, Psychometrics and Statistics. It also includes his ORCID iD link and his location, Netherlands.

university of groningen

Home Profiles Research Units Research output Projects Datasets Prizes ...

 **Tom Bergkamp**  
MSc.  
PhD student, Faculty of Behavioural and Social Sciences, Psychometrics and Statistics  
<https://orcid.org/0000-0001-5522-1247>

Netherlands

**Find a collaborator close to the field**

**They can tell you how things are done,  
the rules of engagement**

# Example 2: Opportunity atlas Microeconomics



# Opportunity atlas

## Question

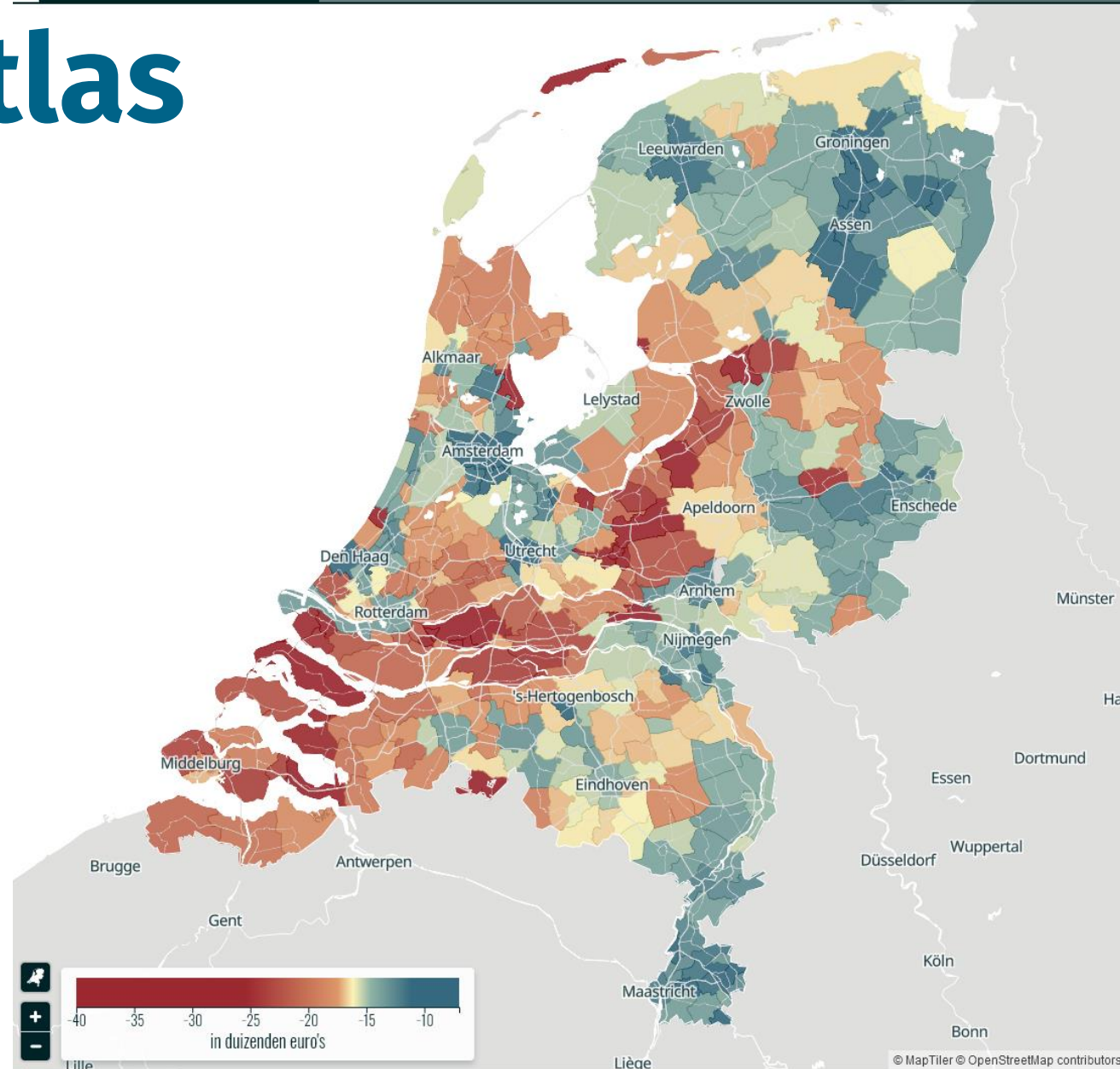
- Throughout NL, how does your parent's income, your migration background, your gender, affect your later life?

## Problem

- There are many categories, many regions, many outcome variables
- This is too slow

## Solution

- Estimation on a supercomputer



**I have a collaborator from the field!**

**Great, but...**

# What went wrong?

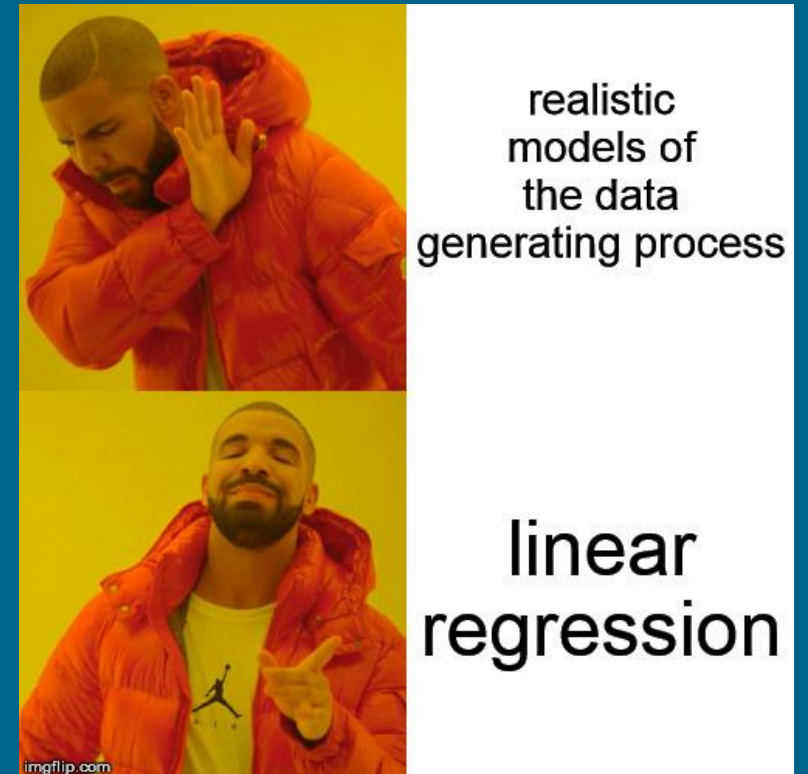
## We spoke totally different languages

An economist from the perspective of a statistician:

- Economists like OLS regression
- Economists like robustness checks
- Economists dislike making assumptions
- Economists write really long papers
- Economists REALLY like OLS regression

NB: there are many good things about the economist's approach to data analysis

<https://github.com/alexpghayes/linear-probability-model>



**Learn the language of your hosts**

**Establish a common way to speak about issues in the project**



# Example 3: Privacy-preserving GLM

Federated machine learning



# Privacy-preserving GLM

## Problem

- Alice & Bob have different variables for the same people
- They are not allowed to combine their dataset for privacy reasons
- They both measure the same outcome variable
- Can we do regression?

## Solution

- Many options exist. We made a new option: blockwise coordinate descent
- Send residuals back & forth, re-estimate based on those residuals until convergence
- Extensions to GLM -> super useful, main focus
- Implemented in software package <https://github.com/vankesteren/privreg>

# What did I do to fit in?

**Worked with collaborators from computer science**

At Maastricht University's Institute of Data Science

**Tried to speak the right language**

Translated all the mentions of “variables” into “features” etc.

Used example data from the UCI Machine Learning repository

**Submitted the paper...**

**Machine  
learning  
people**

**Erik-Jan**



# Reviews



How do authors generalize for the case  $N > 2$  (lot of Alice and Bob)?

The novelty of this work is in my opinion low. The extension to generalized linear models is also rather straightforward.

**Find out what is important and what is trivial to the field**

**What should you focus on in your paper?**



# (Not) Trespassing, summarised

## Be intellectually modest

- Find a collaborator close to the field
- Learn the language of your hosts
- Find out what is important and what is trivial to the field





# (Not) Trespassing, summarised

## Be intellectually modest

- Find a collaborator close to the field
- Learn the language of your hosts
- Find out what is important and what is trivial to the field

**Suggestions?**



# Some experiences

- Trespassing takes time & effort
- You will probably face rejection
- Worthwhile because you will learn a lot

# Epistemic Trespassing

Nathan Ballantyne ✉

*Mind*, Volume 128, Issue 510, April 2019, Pages 367–395, <https://doi.org/10.1093/mind/fzx042>

**Published:** 12 February 2018

“ Cite 🔑 Permissions 🔄 Share ▼

## Abstract

Epistemic trespassers judge matters outside their field of expertise. Trespassing is ubiquitous in this age of interdisciplinary research and recognizing this will require us to be more intellectually modest.

**Thank you!**