

# PETR ŠIMEČEK

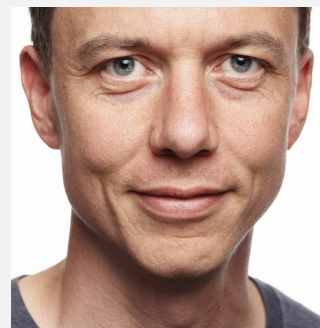
Data Scientist, Bioinformatics Analyst, ML Engineer

## PROFESSIONAL EXPERIENCE

- 2007 | 2009**
  - Biostatistician**  
Institute of Animal Science  
Prague, Czechia
    - Designing experiments
    - Categorical data analysis
    - Mixed-effects models
    - GPS tracking data
- 2007 | 2017**
  - Bioinformatician**  
Institute of Molecular Genetics  
Prague, Czechia
    - Mouse genetics
    - Next generation sequencing
    - Metabolomics
    - Later Head of Bioinformatics Unit
    - IMPC database
- 2013 | 2017**
  - Bioinformatics Analyst**  
The Jackson Laboratory  
Bar Harbor, Maine, USA
    - QTL mapping
    - Mouse diversity outbred
    - Mediation analysis
    - Aging and its effect on proteome
    - R/Shiny & Docker
- 2017 | 2018**
  - Data Scientist**  
Google LLC  
Mountain View, California, USA
    - Time Series: development and maintenance of internal time series forecasting tool
    - Various ad hoc analysis
    - Deep learning applied to time series forecasting
- 2019**
  - Machine Learning Engineer**  
Central European AI Institute (CEAI)  
Brno, Czechia
    - ML model to predict house prices
    - Gradient boosting (XGBoost, LightGBM, CatBoost), neural networks (fast.ai, keras, TF)
    - Amazon EC2, S3, ECS, Elastic Beanstalk, CloudWatch, Apache Airflow

## TEACHING AND SELECTED TALKS

- 2015 | 2017**
  - Introduction to R Language for Beginners.**  
Instructor of Software Carpentry and Software for Scientists,  
<https://crabhi.github.io/2016-10-08-umg/>  
Boston, USA & Prague, Czechia
- 2018 | 2019**
  - Deep Learning: From Zero To Hero in Two Hours.**  
Workshop with intro to deep learning (together with Karla Fejfarova),  
<https://github.com/simecek/from0toheroIn2h>  
Prague, Czechia
- 2019**
  - Statistical vs. Deep Learning Methods for Time Series Forecasting.**  
Recent talk at Machine Learning Meetup,  
<https://youtu.be/mqYwy5RuSQQ>  
Brno, Czechia



## CONTACT INFO

✉ [simecek@gmail.com](mailto:simecek@gmail.com)  
🐙 [github.com/simecek](https://github.com/simecek)  
🐦 [twitter.com/python\\_tip](https://twitter.com/python_tip)  
☎ +420 723-485-497

For more information, please contact me via email.

## SKILLS

Python, gradient boosting (LightGBM, CatBoost), neural networks (PyTorch, TF). Cloud enthusiast.

Fluent in R: Shiny, Rcpp, Tidyverse, RMarkdown. Time series forecasting. QTL Mapping.

Some experience with C++, JavaScript, Bash, LaTeX, Perl, Scala, ...

*This resume was made with the R package [pagedown](#). The source file can be found on my [Github](#).*



## EDUCATION

1998  
|  
2003

### Charles University in Prague

Mgr. (M.Sc.) in Probability Theory and Stochastic Processes (1st prize in the diploma–thesis competition at Department of Probability and Mathematical Statistics in July 2003)

📍 Prague, Czechia

Thesis: On the Minimal Probability of Intersection of Dependent Events

2002

### Vrije Universiteit

Socrates / Erasmus Exchange

📍 Amsterdam, Netherlands

2004  
|  
2005

### Hasselt Universiteit

Master of Science in Biostatistics

📍 Hasselt, Belgium

• AIA Fellowship (one of two annually awarded to Czech students)

• MSc. degree with the great distinction

Thesis: Gene Expression Data Analysis for In Vitro Toxicology

2003  
|  
2007

### Charles University in Prague

Ph.D. in Mathematical Statistics and Probability Theory (thesis summary at <http://bit.ly/2SazFPc>)

📍 Prague, Czechia

Thesis: Independence Models



## SELECTED PUBLICATIONS

See my [Google Scholar profile](#) for the full list of 20+ papers and >750 citations.

2015

### Genetic analysis of substrain divergence in non-obese diabetic (NOD) mice.

G3: Genes, Genomes, Genetics. 2015 May 1;5(5):771-5.

Simecek P, Churchill GA, Yang H, Rowe LB, Herberg L, Serreze DV, Leiter EH.

2016

### Defining the consequences of genetic variation on a proteome-wide scale.

Nature. 2016 Jun;534(7608):500.

Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP.

2017

### High-resolution maps of mouse reference populations.

G3: Genes, Genomes, Genetics. 2017 Oct 1;7(10):3427-34.

Simecek P, Forejt J, Williams RW, Shiroishi T, Takada T, Lu L, Johnson TE, Bennett B, Deschepper CF, Scott-Boyer MP, de Villena FP.