



# Christoph Waldhauser

## Curriculum vitæ

### Education

- 2013–present **PhD Quantitative Methods**, *Department of Methods*, University of Vienna.  
Topic: Text mining, sentiment analysis and topic models in social media
- 2009 **Master of Arts**, *University of Vienna*, Vienna, *With distinction*.
- 2000–2009 **Studies in Statistics, Political Science, Data Engineering, and African Studies**, *University of Vienna & Vienna University of Technology*, Vienna.

### Experience

- 2016–present **Data Scientist**, e-dialog GmbH, Vienna.  
Bootstrapping data science solutions focusing on web analytics, consumer intelligence and workflow automation; development recommendation engines and data visualization solutions.
- 2013–present **Founder & CEO**, KDSS K Data Science Solutions, Vienna.  
Providing business intelligence and data analytical services to clients; focus on spatial statistics, visualization, and interactive data apps.
- 2016 **Chief Data Scientist**, Clozer Ltd., London, UK and Vienna, Austria.  
Developing industry classifications and matching, growth hacking
- Fuzzy industry classifications
    - algorithms for tree traversal and proximity matching
    - Textmining Common Crawl corpus, 145 TB
  - DB design and administration
    - data validation
    - geospatial data
  - Product development

Amerlingstraße 4/14 – A-1060 Vienna

📞 +43 660 5490 623 • ✉ christoph.waldhauser@gmail.com • 🌐 knutur.at  
in christophwaldhauser • 📱 tophcito

- 2014–2016 **Data Scientist**, *Dialog Data*, HEROLD Business Data GmbH, Mödling.  
Development of state of the art methods in CRM, customer segmentation and profiling
- CRM
    - development of classification methodologies
    - PCA based clustering of businesses
    - survival models for customer lifetime value
    - interactive visualizations
    - automated report generation
  - Big Data
    - integration of heterogeneous data sources
    - profiling based on Google Analytics/BigQuery
    - process automation
  - Spatial statistics & location based services
    - propensity modelling based on kNN and krigging
    - GIS applications
- 2012–2014 **Researcher**, *Institute for Statistics and Mathematics*, WU Vienna University of Economics and Business.  
Classification of large data sets from aerial surveying
- Development of methods for automated classification:
    - unsupervised learning: k-means and fuzzy cluster
    - supervised learning: CART decision trees
    - Smoothing: kNN and k-d-trees
  - Prototyping:
    - Package development in R for Linux/Windows
    - Data management (> 10 TB)
  - Presentations at conferences and publications
- 2012–2014 **President**, *Verein StudentInnenkinder daycare center*, Vienna.  
Managing a daycare center with 70 children and 12 employees
- 2010–2012 **eDeveloper**, *Institute for Statistics and Mathematics*, WU Vienna University of Economics and Business.  
Development of interactive learning and teaching aids for Web, Android and iOS
- 2008–2012 **Researcher**, *SORA Institute for Social Research & Consulting*, Vienna.  
All levels of research and consulting: conception, surveying, analysis, interpretation, reporting

## Teaching experience

- 2010–2016 **Lecturer for Statistics**, *Institute for Statistics and Mathematics*, WU Vienna University of Economics and Business.  
Conducting, grading and enhancing WU's introductory course in statistics
- Contents:
    - Inferential statistics
    - Linear models
    - Robust analyses
    - Time series analysis, econometrics
    - Stochastic processes
  - Mode: Lecture, ca. 180 students/term
  - Administration:
    - Led team in designing interactive web-based learning environments (Shiny web apps)
    - Exam development using R/exams

- 2015 **R Workshops**, *Webster Vienna Private University*, Vienna.  
Introductory and advanced R workshops for students and faculty
- Contents:
    - R/RStudio workflow
    - Inferential statistics
    - Visualization
    - Programmatic access to data via APIs
  - Mode: interactive software course (in English)
- 2014 **R Course**, *GIGA German Institute of Global and Area Studies*, Hamburg.  
Introduction to R for Social Sciences and Economics
- Contents:
    - R/RStudio workflow
    - Inferential statistics
    - Cross Sectional Time Series, Panel data
    - Visualization
    - Visualization of geo-referenced data
  - Mode: interactive software course at postgraduate level (in English)
- 2014 **R Courses**, *ECPR Winter School in Methods and Techniques*, Vienna.  
Introduction to R for Social Sciences and Economics
- Contents:
    - R/RStudio workflow
    - Inferential statistics
    - Linear models
    - Decision trees
    - Principal component analysis
    - Factor analysis
    - Missing value treatment
    - Survey weights
    - Visualization
    - Report generation using Markdown
  - Mode: 2 interactive software courses at postgraduate level (in English)
- 2012 **Statistical software**, *SOQUA+ Statistikwochen*, Vienna.  
Managerial implications of statistical package choices
- 2011 **Event history and multi-level analyses**, *SOQUA Program*, Vienna.  
Seminar on advanced methods
- Contents:
    - Data consideration
    - Non-parametric methods
    - Cox' PH-model
    - Mixed-effects models
    - Interpretation
    - Visualization
  - Mode: Small-group course with focus on theory at postgraduate level

## 2010 **Quantitative Methods**, *SOQUA Program*, Vienna.

Introduction to statistics for social sciences with R

- Contents:
  - Working with R
  - Inferential statistics
  - Linear models
  - Decision trees
  - Principal component analysis
  - Factor analysis
  - Missing value treatment
  - Survey weights
- Mode: interactive software course at postgraduate level

---

## Selected Publications

Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Korzeniowska, K., and Wagner G. 2014. Automated Classification of Airborne Laser Scanning Point Clouds. In: S. Koziel, L. Leifsson, and X. Yang, Eds., *Solving Computationally Extensive Engineering Problems: Methods and Applications*. Springer: Berlin Heidelberg, *in press*.

Hochreiter, R. und Waldhauser, C. 2014. Effects of Sampling Methods on Prediction Quality. The Case of Classifying Land Cover Using Decision Trees. *Proceedings of COMPSTAT 2014*, 585–592.

Hochreiter, R. and Waldhauser, C. 2014. Data Mining Cultural Aspects of Social Media Marketing. In: P. Perner, Ed., *Advances in Data Mining, Lecture Notes in Computer Science*. Springer: Berlin Heidelberg, 130–143.

Hochreiter, R. and Waldhauser, C. 2013. A Stochastic Simulation of the Decision to Retweet. In: P. Perny, M. Pirlot and A. Tsoukiàs, Eds., *Algorithmic Decision Theory. Lecture Notes in Computer Science*. Springer: Berlin Heidelberg, 221–229.

Otepka, J., Ghuffar, S., Waldhauser, C., Hochreiter, R., and Pfeifer, N. 2013. Georeferenced Point Clouds: A Survey of Features and Point Cloud Management. *ISPRS International Journal of Geo-Information* 2, 4, 1038–1065.

Hochreiter, R. and Waldhauser, C. 2013. Solving Dynamic Optimisation Problems with Revolutionary Algorithms. *International Journal of Innovative Computing and Applications* 5, 3, 143–151.

Tatzl, G. and Waldhauser, C. 2010. Aggregating Opinions: Explorations into Graphs and Media Content Analysis. *Proceedings of textGraphs-5, Association for Computational Linguistics*, 93–97.

---

## Methods

### Statistics

- Inferential statistics
- Principal component analysis
- Geostatistics
- Regression (OLS, logistic, Poisson)
- Event history & survival analysis
- Complex samples
- Cluster & factor analysis
- Multidimensional scaling
- Mixed-effects models
- Network analysis

### Data mining & Optimization

- Text mining
- Genetic algorithms
- Web scraping
- Machine learning

Amerlingstraße 4/14 – A-1060 Vienna

☎ +43 660 5490 623 • ✉ christoph.waldhauser@gmail.com • 🌐 knutur.at

in christophwaldhauser • 🐙 tophcito

---

## Software

### Programming

- R Interactive analyses, scripting, package development, web apps with Shiny, parallelization
- Scala Process automation, data manipulation, machine learning
- Hadoop Data mangement, data warehousing, MapReduce
- Spark Machine learning, stream processing, graph processing
- SPSS Interactive analyses, scripting
- SAS Basic analyses, data mining
- VBA Process automation, user interfaces
- Bash Scripting
- SQL Database and query design
- Git Application, collaboration, administration
- HTML Websites in HTML4 and HTML5 using CSS and JS; Bootstrap
- Javascript Full stack web apps using Angular.js, Node.js, and MongoDB

### Applications

- Office Proficient in text processing, spreadsheets, presentations
- $\LaTeX$  Creating of complex documents; class authoring
- Emacs ESS, Auc $\TeX$ , NoWeb for embedding R code in  $\LaTeX$  and Markdown
- Linux Using Gnome desktop and CLI
- GIS Preparing cartographic material and visualizing data

### Administration

- Linux Administration of Ubuntu, Gentoo instances; Virtualization, Docker
- Web server Administration of Apache, LightHTTP, Nginx

---

## Languages

|         |          |         |       |
|---------|----------|---------|-------|
| German  | Native   | Hausaa  | Basic |
| English | Native   | Polish  | Basic |
| French  | Advanced | Spanish | Basic |

---

## Memberships & Reviewing

- ACM Association for Computing Machinery
- EPSA European Political Science Association
- ÖSG Austrian Statistical Society
- IASE-ISLP International Association for Statistical Education, *ISLP Country coordinator*
- TISE Technology Innovation in Statistics Education, *Reviewer*

Amerlingstraße 4/14 – A-1060 Vienna

☎ +43 660 5490 623 • ✉ christoph.waldhauser@gmail.com • 🌐 knutur.at  
in christophwaldhauser • 📺 tophcito