Andee Kaplan

CONTACT INFORMATION

Department of Statistical Science
P.O. Box 90251
Durham, NC 27708-0251

Phone: (832) 526-7947
E-mail: andrea.kaplan@duke.edu
WWW: `andeekaplan.com`

EDUCATION

Iowa State University (ISU), Ames, Iowa USA

Ph.D., Statistics, August 2017

- Dissertation topic: "On advancing MCMC-based methods for Markovian data structures with applications to deep learning, simulation, and resampling"
- Advisors: Daniel Nordman, Stephen Vardeman

M.S., Statistics, May 2014

- Research topic: "gravicom - a web-based tool for community detection in networks"
- Advisors: Heike Hofmann, Daniel Nordman

The University of Texas, Austin, Texas USA

M.A., Mathematics, December 2010

- Research topic: "An Overview of Multilevel Regression"
- Advisors: Martha Smith, John Luecke

B.S., Mathematics, May 2006

- Elements of Computing Certificate

PROFESSIONAL APPOINTMENTS

Duke University, Durham, NC USA

Postdoctoral Associate, Department of Statistical Science, August 2017 - Present

- Advisor: Rebecca C. Steorts

RESEARCH INTERESTS

Statistical Machine Learning, Bayesian Statistics, Computational Statistics, Record Linkage, Differential Privacy, Markov chain Monte Carlo and Spatial Resampling, Interactive Statistical Graphics, Reproducible Research, Social Statistics, Human Rights

HONORS AND AWARDS

ISU Department of Statistics George W. Snedecor Award 2015
This award honors the founder and first director of the Statistical Laboratory, George W. Snedecor, and is awarded to the most outstanding PhD candidate in the Department of Statistics.

American Statistical Association Computing Section Student Paper Award 2015
Awarded for paper "Introductory statistics with intRo."

ISU Department of Statistics Holly and Beth Fryer Scholarship 2014
Criteria for this scholarship include grades received in Statistics and related courses, performance in assistantship duties and other information that indicates a high likelihood that the student will make contributions to the Statistics profession throughout their career.

Women in Statistics Conference Poster Award 2014
Awarded for poster "Mathematical Self-Efficacy of Incoming Students at a Large Public University."

American Statistical Association Data Exposition 1^{st} *Place* 2013
Awarded for poster "Putting Down Roots: A Graphical Exploration of Community Attachment."

ISU Department of Statistics Vera David Fellowship 2013

This fellowship, is given to a female student who has just completed her first year of graduate studies. The scholarship is awarded on the basis of academic achievement during the students first year.

NSF Research Experiences for Undergraduates 2006
Participant in Extensible Undergraduate Research in Communications Applications, in the areas of Communications, Networks and Systems.

University of Texas at Austin Honors and Dean's List 2005 & 2006

Other Awards

IMS New Researchers Conference Travel Award 2018
IMA Frontiers in Forecasting Travel Award 2018
Conference on Data Analysis Poster Grant 2016
rOpenSci Unconf Travel Funding 2015
Fields Institute Workshop on Visualization for Big Data Travel Award 2015
Women in Statistics Conference Travel Award 2014
ASA Joint Statistical Meetings Special Student Funding Travel Award 2014

Refereed Journal Publications

[1] Hare, E. and **Kaplan, A.** "Designing Modular Software: A Case Study in Introductory Statistics". *Journal of Computational and Graphical Statistics* 26.3 (2017), pp. 493–500. DOI: 10.1080/10618600.2016.1276839.

[2] **Kaplan, A.**, Hofmann, H., and Nordman, D. "An interactive graphical method for community detection in network data". *Computational Statistics* 32.2 (2017), pp. 535–557. ISSN: 1613-9658. DOI: 10.1007/s00180-016-0663-5.

[3] **Kaplan, A.**, Hare, E., Hofmann, H., and Cook, D. "Can you buy a president? Politics after the Tillman Act". *Chance* 27.1 (2014), pp. 20–30.

Preprints

[4] **Kaplan, A.**, Kaiser, M., Lahiri, S., and Nordman, D. "Simulating Markov random fields with a conclique-based Gibbs sampler". *arXiv preprint arXiv:1808.04739* (in review). URL: https://arxiv.org/abs/1808.04739.

[5] **Kaplan, A.**, Nordman, D., and Vardeman, S. "On the instability and degeneracy of deep learning models". *arXiv preprint arXiv:1612.01159* (in review). URL: https://arxiv.org/abs/1612.01159.

[6] **Kaplan, A.** and Hare, E. "Putting Down Roots: A Graphical Exploration of Community Attachment". *arXiv preprint arXiv:1608.04843* (in review). URL: http://arxiv.org/abs/1608.04843.

[7] **Kaplan, A.**, Betancourt, B., and Steorts, R. C. "Posterior Prototyping: Bridging the Gap between Record Linkage and Regression" (to be submitted).

[8] **Kaplan, A.**, Nordman, D., and Vardeman, S. "Properties and Bayesian fitting of restricted Boltzmann machines". *arXiv preprint arXiv:1612.01158* (to be submitted). URL: https://arxiv.org/abs/1612.01158.

Refereed Conference Preceedings

[9] Mouzon, I., Genschel, U., Nguyen, X. H., **Kaplan, A.**, Carriquiry, A., and Mann, C. "A Cluster Analysis of STEM Gender Differences". *Proceedings of the 18th Annual Conference on Research in Undergraduate Mathematics Education.* The SIGMAA on Research in Undergraduate Mathematics Education. 2015, pp. 793–800.

Unrefereed Conference Preceedings

[10] Hofmann, H., Cook, D., **Kaplan, A.**, Hare, E., Leos-Barajas, V., Sievert, C., and Tyner, S. "On the move at DinoFun world". *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on.* IEEE. 2015, pp. 159–160.

[11] Hofmann, H., Cook, D., **Kaplan, A.**, Hare, E., Leos-Barajas, V., Sievert, C., and Tyner, S. "Visualizing communication patterns at DinoFun World". *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*. IEEE. 2015, pp. 161–162.

[12] Genschel, U., **Kaplan, A.**, Carriquiry, A., Johnston, E., Kliemann, W., Koehler, K., Mouzon, I., and Nguyen, H. "Statistical and Mathematical Self-Efficacy of Incoming Students at a Large Public University". *Sustainability in Statistics Education, Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA*. 2014.

Software

[1] `representr`, Create Representative Records Post-Record Linkage, with Brenda Betancourt and Rebecca C. Steorts [7]. (Forthcoming)

[2] `sblink`, Scalable Empirical Bayes Entity Resolution, with Neil Marchant, Ben Rubenstein, and Rebecca C. Steorts. (Forthcoming)

[3] `conclique`, Gibbs Sampling for Spatial Data and Concliques, with Daniel Nordman, Mark Kaiser, and Soumendra Lahiri [4]. (`https://github.com/andeek/conclique`)

[4] `intro`, Download and Run the intRo Statistical Software, with Eric Hare [1]. (`https://github.com/gammarama/intro`)

[5] `forestr`, Random Forests with a User Created Splitting Criteria. (`https://github.com/andeek/forestr`)

Invited Talks

*Denotes presenter.

[1] **Kaplan, A.*** and Steorts, R. C. "Population Sized Graphical Record Linkage". *Invited Session*. International Conference on Advances in Interdisciplinary Statistics and Combinatorics. Greensboro, NC, USA, Oct. 2018 (Upcoming).

[2] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A Fast Sampler for Data Simulation from Markov Random Fields". *Invited Session*. The 28th Annual Conference of the International Envirometrics Society. Guanajuato, Gto., Mexico, July 2018.

[3] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Keynote*. The 1st Midwest Statistical Machine Learning Colloquium. Ames, IA, USA, May 2018.

[4] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A simple, fast sampler for simulating spatial data and other Markovian data structures". *Statistics Seminar*. Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2018.

[5] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A simple, fast sampler for simulating spatial data and other Markovian data structures". *Statistics Seminar*. North Carolina State University, Department of Statistics. Raleigh, NC, USA, Nov. 2017.

[6] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A simple, fast sampler for simulating spatial data and other Markovian data structures". *Computer Science Seminar*. Centro de Investigación en Matemáticas. Guanajuato, Gto., Mexico, Oct. 2017.

[7] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A simple, fast sampler for simulating spatial data and other Markovian data structures". *Colloquium*. Duke University, Department of Statistical Science. Durham, NC, USA, Apr. 2017.

[8] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. Michigan State University, Department of Statistics and Probability. East Lansing, MI, USA, Feb. 2017.

[9] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. University of Illinois, Urbana-Champaign, Department of Statistics. Chapaign, IL, USA, Feb. 2017.

[10] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. University of Florida, Department of Statistics. Gainesville, FL, USA, Feb. 2017.

[11] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. Rice University, Department of Statistics. Houston, TX, USA, Jan. 2017.

[12] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. University of Virginia, Department of Statistics. Charlottesville, VA, USA, Jan. 2017.

[13] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. Arizona State University, School of Mathematical and Statistical Sciences. Tempe, AZ, USA, Jan. 2017.

[14] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Model matters with restricted Boltzmann Machines". *Colloquium*. University of Massachussetts, Amherst, Department of Mathematics & Statistics. Amherst, MA, USA, Dec. 2016.

[15] **Kaplan, A.*** and Nordman, D. "Goodness-of-Fit Tests for Spatial Markov Random Fields". *8th International Conference of the ERCIM WG on Computational and Methodological Statistics*. Organized Invited Session. European Research Consortium for Informatics and Mathematics. London, UK, Dec. 2015.

[16] Hare, E.* and **Kaplan, A.***. "intRo: Statistical Analysis Software for Teaching". *Departmental Seminar*. Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2015.

[17] **Kaplan, A.*** and Hare, E.*. "Putting Down Roots: A Graphical Exploration of Community Attachment". *Census Data Visualization Seminar*. U.S. Census Bureau. Washington, D.C., USA, Sept. 2013.

Invited Short Courses

Some of Record Linkage, US Census Bureau & Centro de Investigación de Matemáticas, A. C.

Co-instructor
Designed and facilitated a workshop that provides a broad introduction to multiple methods for performing record linkage and deduplification, including blocking and evaluation topics (`https://resteorts.github.io/record-linkage-tutorial/`).

Machine Learning Day, Duke University

Co-organizer
Helped organize Duke's first Machine Learning Day for undergraduates and participated as a panel speaker.

D3 Workshop, NORC at the University of Chicago

Co-instructor
Designed and presented a workshop for NORC employees on the JavaScript plotting library D3 (`http://andeekaplan.com/d3workshop/`).

Week of R, Iowa State University

Co-instructor
Student run R workshop within the Department of Statistics covering the basics of R, plotting, data manipulation, scraping web data, and Shiny (`http://heike.github.io/rwrks/`).

Contributed Talks

[1] **Kaplan, A.*** and Steorts, R. C. "Counting Casualties in the Syrian Civil War with Bayesian Record Linkage". *Joint Statistical Meetings*. Topic Contributed Paper. ASA. Vancouver, BC, Canada, July 2018.

[2] **Kaplan, A.*** and Steorts, R. C. "Population Sized Graphical Record Linkage". *ISBA World Meeting*. Contributed Member Paper. ISBA. Edinburgh, UK, June 2018.

[3] **Kaplan, A.***, Kaiser, M., Lahiri, S., and Nordman, D. "A Simple, Fast Sampler for Simulating Spatial Data and Other Markovian Data Structures". *Joint Statistical Meetings*. Contributed Paper. ASA. Baltimore, MD, USA, Aug. 2017.

[4] **Kaplan, A.***, Nordman, D., and Vardeman, S. "An exposition on the propriety of restricted Boltzmann machines". *Joint Statistical Meetings*. Contributed Paper. ASA. Chicago, IL, USA, July 2016.

[5] **Kaplan, A.*** and Hare, E. "Introducing Statistics with intRo". *The R User Conference*. Contributed Paper. Stanford, CA, USA, June 2016.

[6] **Kaplan, A.*** and Hare, E. "Introducing Statistics with intRo". *Joint Statistical Meetings*. Topic Contributed Paper. ASA. Seattle, WA, USA, Aug. 2015.

[7] Hare, E.* and **Kaplan, A.** "IntRo: Software for Introductory Statistics". *Joint Statistical Meetings*. Contributed Paper. ASA. Seattle, WA, USA, Aug. 2015.

[8] **Kaplan, A.***, Hofmann, H., and Nordman, D. "gravicom - A web-based tool for community detection in networks". *Joint Statistical Meetings*. Contributed Paper. ASA. Boston, MA, USA, Aug. 2014.

[9] Mouzon, I.*, Genschel, U., Carriquiry, A., Nguyen, X. H., **Kaplan, A.**, Johnston, E., Kliemann, W., and Koehler, K. "Early College Performance, Gender, and Other Factors Influencing Continuation in STEM Fields". *Joint Statistical Meetings*. Contributed Paper. ASA. Boston, MA, USA, Aug. 2014.

[10] Genschel, U.*, **Kaplan, A.**, Carriquiry, A., Johnston, E., Kliemann, W., Koehler, K., and Mouzon, I. "Statistical and Mathematical Self-Efficacy of Incoming Students at a Large Public University". *Joint Statistical Meetings*. Contributed Paper. ASA. Boston, MA, USA, Aug. 2014.

[11] Genschel, U.*, **Kaplan, A.**, Carriquiry, A., Johnston, E., Kliemann, W., Koehler, K., and Mouzon, I. "Statistical and Mathematical Self-Efficacy of Incoming Students at a Large Public University". *The Ninth International Conference on Teaching Statistics*. Contributed Paper. Flagstaff, AZ, USA, July 2014.

Contributed Posters

[1] **Kaplan, A.*** and Steorts, R. C. "Population Sized Record Linkage". *New Researchers Conference*. Contributed Poster. Institute of Mathematical Statistics. Vancouver, BC, Canada, July 2018.

[2] **Kaplan, A.***, Nordman, D., and Vardeman, S. "Properties and Bayesian fitting of restricted Boltzmann machines". *Frontiers in Forecasting*. Contributed Poster. Institute for Mathematics and its Application, University of Minnesota. Minneapolis, MN, USA, Feb. 2018.

[3] **Kaplan, A.***, Nordman, D., and Vardeman, S. "An Exposition on the Propriety of Restricted Boltzmann Machines". *Conference on Data Analysis*. Contributed Poster. Los Alamos National Laboratory. Santa Fe, NM, USA, Mar. 2016.

[4] **Kaplan, A.**, Lin, Y., and Mulrow, E.*. "Visualizing Linked Data Sources for the National Children's Study". *Conference on Statistical Practice*. Contributed Poster. ASA. San Diego, CA, USA, Feb. 2016.

[5] **Kaplan, A.***, Genschel, U., Carriquiry, A., Johnston, E., Kliemann, W., Koehler, K., Mouzon, I., and Nguyen, X. H. "Mathematical Self-Efficacy of Incoming Students at a Large Public University". *Women in Statistics Conference*. Student Poster. ASA. Cary, NC, USA, May 2014.

[6] **Kaplan, A.*** and Hare, E. "Putting Down Roots: A Graphical Exploration of Community Attachment". *Joint Statistical Meetings*. Data Exposition Poster. ASA. Montreal, QC, Canada, Aug. 2013.

[7] Hare, E.* and **Kaplan, A.** "Can you buy a president? Politics after the Tillman Act". *Joint Statistical Meetings*. Contributed Poster. ASA. Montreal, QC, Canada, Aug. 2013.

Teaching Experience

Duke University, Durham, North Carolina USA

Guest Lecturer **Fall 2017 & Spring 2018**
Presented material for Data Mining and Machine Learning (Undergraduate level) and Bayesian Methods and Modern Statistics (Masters level) courses.

Iowa State University, Ames, Iowa USA

Instructor **Summer 2017**
Course instructor for Engineering Statistics (`http://andeekaplan.com/stat305`). Statistics for engineering problem solving, covering principles of engineering data collection, descriptive statistics, elementary probability distributions, principles of experimentation, confidence intervals and significance tests, one- and two-sample studies, regression analysis, and statistical software.

Instructor **Spring 2017**
Course instructor for Engineering Statistics (`http://andeekaplan.com/stat305/spring2017`). Statistics for engineering problem solving, covering principles of engineering data collection, descriptive statistics, elementary probability distributions, principles of experimentation, confidence intervals and significance tests, one- and two-sample studies, regression analysis, and statistical software.

Instructor **Fall 2016**
Course instructor for Data Stewardship for Earth Systems Scientists (`http://agron590-ISU.github.io`). Co-designed and delivered course material focusing on fundamental data skills required for successful, collaborative, and reproducible research within the context of plant, soil, and atmospheric sciences.

Guest Lecturer **Spring 2015**
Presented material for one week of Advanced Probability Theory (PhD level) and one week of Time Series Analysis (Masters level) courses.

Teaching Assistant **Fall 2014**
Lab instructor for Introduction to Business Statistics II. Responsible for running lab section and grading exams. Approximately 30 students.

Instructor **Spring 2013**
Course instructor for Introduction to Business Statistics. Responsible for lecture and the development of course notes, homework assignments, and exams. Approximately 80 students.

Grading Coordinator **Fall 2012**
Coordinated a team of seven undergraduates to grade for Introduction to Business Statistics. Duties included creating a rubric each week and ensuring consistent grading across all sections.

Research Students

Undergraduate Students
Ritika Bharati, Undergraduate research project, Spring 2018 - Present
Srini Sunil, Undergraduate research on blocking methods for record linkage, Fall 2017 - Present

Research Experience

Duke University, Durham, North Carolina USA

Postdoctoral Associate **August, 2017 - Present**
Scalable Record Linkage through the Microclustering Property, funded by NSF, with Steorts, R. C.

Iowa State University, Ames, Iowa USA

Graduate Research Assistant **May, 2015 - August, 2016**
Nonparametric Likelihood Enhancements for Dependent Data, funded by NSF, with Nordman, D.

Graduate Research Assistant **May, 2013 - August, 2015**
Exploring the STEM Gender Gap: Introductory College Mathematics and Statistics Instruction and its Association with Self-Efficacy, funded by NSF, with Genschel, U., Carriquiry, A., Kliemann, W., Johnston, E., Koehler, K., Nguyen, X., Mouzon, I.

The University of Texas, Austin, Texas USA

Graduate Research Assistant **April, 2011 - July, 2012**
Worked with Prof. Tasha Beretvas to assess use of Bayesian methods to estimate a two-level cross-classified random effects model with correlated Level 2 residuals.

Summer REU **June, 2006 - August, 2006**
Participant in Extensible Undergraduate Research in Communications Applications, a summer REU in the Department of Electrical Engineering sponsored by the National Science Foundation for research in the areas of Communications, Networks and Systems. Worked on optimization of binary erasure channels project with Prof. Sriram Vishwanath.

Undergraduate Research Assistant **June, 2005 - March, 2006**
Worked with Prof. Tasha Beretvas to identify different methods of effect size estimation and evaluate the advantages and disadvantages of each method. Developed a Java program for data entry.

Applied Research Laboratories, Austin, Texas USA

Graduate Research Assistant **August, 2006 - February, 2007**
Research with a concentration in active and passive data fusion. Programmed graphical work in signal processing using Matlab.

Industry Experience

NORC at the University of Chicago, Chicago, Illinois USA

Graduate Research Assistant **May, 2015 - August, 2015**
Summer internship with Statistics and Methodology Department. Primarily developed web-based interactive graphics using JavaScript library D3 to explore data linkage for extant sources. Additionally, performed data munging and low level data analysis as well as report generation for third-party clients in R and R Markdown.

Banks Information Group, Austin, Texas USA

Independent Contractor **July, 2011 - July, 2012**
Developed web-based environmental GIS reporting tool for Phase I reporting. Additionally, developed and implemented a web-based tool to convert images to PDF files programatically. Utilized ESRI ArcSDE and .NET/C#, as well as Microsoft SQL Server for development.

Sense Corp, Austin, Texas USA

Consultant **July, 2009 - July, 2011**
Designed monthly subscriber activity forecasts across key markets at a top-five broadband telecommunications company using ARIMA time series modeling in R. Developed an automated process and custom web application designed for Business Intelligence (BI) Group use to run forecasts and analyze long-term trends in activity. Analyzed the effects of explanatory factors on elevated truck rolls and presented findings to corporate leadership.

Gelb Consulting Group, Houston, Texas USA

Intern Analyst **June, 2008 - August, 2008**

Internal consultant focusing on revising several standard operating procedures, including regression and factor analysis.

Service and Leadership

Statistics in the Community (StatCom)

Network Outreach Coordinator **2015 - 2017**
Executive Committee, ISU Chapter **2013 - 2017**

Iowa State University

Student representative to departmental faculty meetings **2015 - 2017**
STATers, social organization for graduate students, Vice President **2013 - 2014**
Graduate and Professional Student Senate, Senator **2012 - 2013**

Reviewer

AIStats, *American Political Science Review*, *Environmental and Ecological Statistics*, *Journal of the American Statistical Association*, *PLOS One*

Conferences

Contributed Session Chair, Joint Statistical Meetings, Vancouver, BC, Canada **August 2018**

Professional Affiliations

American Statistical Association (ASA), Institute of Mathematical Statistics (IMS)

Computing

- Mathematical and Statistical Computing: R, Rcpp, Shiny, Julia, JAGS, BUGS, SAS, Matlab; some experience with Python and SPSS.
- Other Languages: C++, JavaScript (and D3), Java, SQL, C#, .NET, HTML5, CSS3, Markdown.
- Content Management: Git