Andee Kaplan

CONTACT INFORMATION

Department of Statistics
Colorado State University
Fort Collins, CO 80523-1877

E-mail: andee.kaplan@colostate.edu
WWW: `andeekaplan.com`
ORCID: 0000-0002-2940-889X

PROFESSIONAL APPOINTMENTS

Colorado State University, Fort Collins, CO USA

Assistant Professor, Department of Statistics, August 2019 - Present

Duke University, Durham, NC USA

Postdoctoral Associate, Department of Statistical Science, August 2017 - July 2019

Advisor: Rebecca C. Steorts

EDUCATION

Iowa State University (ISU), Ames, Iowa USA

Ph.D., Statistics, August 2017

Dissertation title: "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 title: "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 title: "An Overview of Multilevel Regression"
Advisors: Martha Smith, John Luecke

B.S., Mathematics, May 2006

Elements of Computing Certificate

RESEARCH INTERESTS

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

HONORS AND AWARDS

Selected for "The Best of JCGS – Invited Papers" Session at JSM 2019
"Designing Modular Software: A Case Study in Introductory Statistics" with E. Hare selected for inclusion in the session.

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 student's 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

ISBA Bayes Comp Conference Travel Award 2020
IMS New Researchers Conference Travel Award 2018 & 2019
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

Grants & Contracts

As PI: *Streaming Record Linkage for Online Data Deduplication* (North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan and Betancourt ($130,351.57; CSU: $90,623.57). January 2021 - December 2021 (Pending).

As PI: *Streaming Record Linkage for Online Data Deduplication* (North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan and Betancourt ($132,397.90; CSU: $92,032.90). January 2020 - December 2020.

As PI: *An Extensible Model for Deduplication of the GDELT Events Database* (North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan ($23,995.51). August 2019 - December 2019.

As Co-PI: *Posterior Prototyping: Bridging the Gap between Record Linkage and Regression* (North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan, Betancourt, and Steorts ($96,897). January 2019 - December 2019.

Refereed Journal Publications

[1] Marchant, N. G., **Kaplan, A.**, Elazar, D. N., Rubinstein, B. I. P., and Steorts, R. C. "d-blink: Distributed End-to-End Bayesian Entity Resolution". *Journal of Computational and Graphical Statistics* 0.ja (2020), pp. 1–42.

[2] **Kaplan, A.**, Kaiser, M. S., Lahiri, S. N., and Nordman, D. J. "Simulating Markov Random Fields With a Conclique-Based Gibbs Sampler". *Journal of Computational and Graphical Statistics* 29.2 (2020), pp. 286–296.

[3] **Kaplan, A.**, Nordman, D., and Vardeman, S. "On the S-instability and degeneracy of discrete deep learning models". *Information and Inference: A Journal of the IMA* 9.3 (2020), pp. 627–655.

[4] **Kaplan, A.** and Hare, E. "Putting Down Roots: A Graphical Exploration of Community Attachment". *Computational Statistics* 34.4 (2019), pp. 1449–1464.

[5] **Kaplan, A.**, Nordman, D., and Vardeman, S. "Properties and Bayesian fitting of restricted Boltzmann machines". *Statistical Analysis and Data Mining: The ASA Data Science Journal* 12.1 (2019), pp. 23–38.

[6] 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.

[7] **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.

[8] **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

[9] Lu, X., Hooten, M., **Kaplan, A.**, Womble, J., and Bower, M. "Improving wildlife population inference using aerial imagery and entity resolution" (2020+).

[10] **Kaplan, A.**, Betancourt, B., and Steorts, R. C. "Entity Resolution and the Downstream Task: A Case Study of North Carolina Voter Registration Records". *arXiv preprint arXiv:1810.01538* (2020+). URL: https://arxiv.org/abs/1810.01538.

Refereed Conference Preceedings

[11] 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

[12] 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.

[13] 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.

[14] 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 [10]. (`https://github.com/cleanzr/representr`)

[2] `dblink`, Scalable Empirical Bayes Entity Resolution, with Neil Marchant, Ben Rubenstein, and Rebecca C. Steorts [1]. (`https://github.com/cleanzr/dblink`)

[3] `conclique`, Gibbs Sampling for Spatial Data and Concliques, with Daniel Nordman, Mark Kaiser, and Soumendra Lahiri [2]. (`https://github.com/andeek/conclique`)

[4] `intro`, Download and Run the intRo Statistical Software, with Eric Hare [6]. (`https://github.com/gammarama/intro`)

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

Invited Talks

*Denotes presenter.

[1] **Kaplan, A.***, Betancourt, B., and Steorts, R. C. "Bayesian Canonicalization of Voter Registration Files". *Symposium on Data Science and Statistics*. Invited Session. ASA. Pittsburgh, PA, USA (Remote via Zoom), June 2020.

[2] **Kaplan, A.***, Betancourt, B., and Steorts, R. C. "Life after record linkage: Tackling the downstream task with error propagation". *Bayes Comp*. Invited Session. ISBA. Gainesville, FL, USA, Jan. 2020.

[3] **Kaplan, A.***, Betancourt, B., and Steorts, R. C. "Posterior Prototyping for Bayesian Entity Resolution". *12th International Conference of the ERCIM WG on Computational and Methodological Statistics*. Organized Invited Session. European Research Consortium for Informatics and Mathematics. London, UK, Dec. 2019.

[4] **Kaplan, A.*** and Hare, E. "Designing Modular Software: a Case Study in Introductory Statistics". *Joint Statistical Meetings*. Invited Session. ASA. Denver, CO, USA, July 2019.

[5] **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.

[6] **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.

[7] **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.

[8] **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.

Departmental Seminars

[1] Iowa State University, Graphics Group. Ames, IA, USA (Remote via Zoom), Nov. 2020.

[2] University of Toronto, Department of Statistics. Toronto, ON, Canada (Remote via Zoom), Nov. 2020.

[3] New York University, Center for Practice, Research at the Intersection of Information, Society, and Methodology. New York, NY, USA (Remote via Zoom), Oct. 2020.

[4] The University of Wisconsin, Systems, Information, Learning and Optimization (SILO) Research Group. Madison, WI, USA, Nov. 2019.

[5] Duke University, Department of Computer Science. Durham, NC, USA, Apr. 2019.

[6] The University of California Irvine, Department of Statistics. Irvine, CA, USA, Feb. 2019.

[7] The University of Kentucky, Department of Statistics. Lexington, KY, USA, Feb. 2019.

[8] The Pennsylvania State University, Department of Statistics. State College, PA, USA, Feb. 2019.

[9] Virginia Tech University, Department of Statistics. Blacksburg, VA, USA, Feb. 2019.

[10] The University of Wisconsin, Department of Statistics. Madison, WI, USA, Feb. 2019.

[11] Colorado State University, Department of Statistics. Fort Collins, CO, USA, Jan. 2019.

[12] The University of Texas at Austin, Department of Statistics and Data Science. Austin, TX, USA, Jan. 2019.

[13] Stanford University, Department of Statistics. Stanford, CA, USA, Jan. 2019.

[14] Monash University, Department of Econometrics and Business Statistics. Melbourne, VIC, Australia, Jan. 2019.

[15] Texas A&M University, Department of Statistics. College Station, TX, USA, Nov. 2018.

[16] North Carolina State University, Department of Statistics. Raleigh, NC, USA, Oct. 2018.

[17] Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2018.

[18] North Carolina State University, Department of Statistics. Raleigh, NC, USA, Nov. 2017.

[19] Centro de Investigación en Matemáticas. Guanajuato, Gto., Mexico, Oct. 2017.

[20] Duke University, Department of Statistical Science. Durham, NC, USA, Apr. 2017.

[21] Michigan State University, Department of Statistics and Probability. East Lansing, MI, USA, Feb. 2017.

[22] University of Illinois, Urbana-Champaign, Department of Statistics. Chapaign, IL, USA, Feb. 2017.

[23] University of Florida, Department of Statistics. Gainesville, FL, USA, Feb. 2017.

[24] Rice University, Department of Statistics. Houston, TX, USA, Jan. 2017.

[25] University of Virginia, Department of Statistics. Charlottesville, VA, USA, Jan. 2017.

[26] Arizona State University, School of Mathematical and Statistical Sciences. Tempe, AZ, USA, Jan. 2017.

[27] University of Massachussetts, Amherst, Department of Mathematics & Statistics. Amherst, MA, USA, Dec. 2016.

[28] Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2015.

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.***, Betancourt, B., and Steorts, R. C. "Life After Record Linkage: Tackling the Downstream Task with Error Propagation". *New Researchers Conference*. Contributed Poster. Institute of Mathematical Statistics. Fort Collins, CO, USA, July 2019.

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

[3] **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.

[4] **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.

[5] **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.

[6] **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.

[7] **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.

[8] 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

CSU	DSCI 445	Statistical Machine Learning	F20
CSU	STAT 730	Advanced Theory of Statistics I	S20
CSU	STAT 400	Statistical Computing	F19, F20
ISU	STAT 305	Engineering Statistics	S17, SU17
ISU	AGRON 590DS	Data Stewardship for Earth Systems Scientists	F16
ISU	STAT 226	Introduction to Business Statistics	S13

Student Advising

PhD Students

Ian Taylor, Co-advising with Bailey Fosdick

Graduate Research Assistants

Ian Taylor, GRA — Spring 2020 - Fall 2020

Casey Schafer, GRA — Fall 2019

Undergraduate Students

Boston Lee (Colorado State University), Honors Option DSCI445 — Fall 2020

Olivia Beck (Colorado State University), Honors Option STAT400 — Fall 2019

Ritika Bharati (Duke University), Undergraduate research — Spring 2018 - Spring 2019

Srini Sunil (Duke University), Undergraduate research — Fall 2017 - Spring 2019

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

Colorado State University

Co-organize Statistical Learning and Data Science Journal Club Fall 2020 - Present
Admissions Committee Fall 2020 - Spring 2021
Data Science Research Initiative Steering Committee Fall 2020 - Spring 2021
Data Science Committee Fall 2019 - Spring 2021
Seminar Co-organizer Fall 2019 - Spring 2020

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*, *Annals of Applied Statistics*, *Environmental and Ecological Statistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society, Section A*, *Journal of Survey Statistics and Methodology*, *PLOS One*, *Statistics and Probability Letters*

NSF Panelist (DMS)

Conferences

Contributed Session Organizer, Joint Statistical Meetings, Philadelphia, PA, USA August 2020
Co-organizer, Symposium for Data Science & Statistics, Pittsburgh, PA, USA June 2020
Invited Session Chair, Joint Statistical Meetings, Denver, CO, USA August 2019
Contributed Session Organizer, Joint Statistical Meetings, Vancouver, BC, Canada August 2018

PROFESSIONAL AFFILIATIONS

American Statistical Association (ASA), International Society for Bayesian Analysis (ISBA)

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