Andee Kaplan

CONTACT Information Department of Statistics *E-mail:* andee.kaplan@colostate.edu Colorado State University *WWW:* https://andeekaplan.com Fort Collins, CO 80523-1877 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: "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

Advisors: Heike Hofmann, Daniel Nordman

The University of Texas, Austin, Texas USA

M.A., Mathematics December 2010

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

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 Awarded for paper "Introductory statistics with intRo."

2015

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 1st 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

REFEREED JOURNAL PUBLICATIONS

- * Denotes student author.
- [1] **Kaplan, A.** and Bien, J. "Interactive Exploration of Large Dendrograms with Prototypes". The American Statistician 0.0 (2022), pp. 1–11.
- [2] **Kaplan, A.**, Betancourt, B., and Steorts, R. C. "A Practical Approach to Proper Inference with Linked Data". *The American Statistician* 76.4 (2022), pp. 384–393.
- [3] Lu, X.*, Hooten, M., **Kaplan, A.**, Womble, J., and Bower, M. "Improving Wildlife Population Inference Using Aerial Imagery and Entity Resolution". *Journal of Agricultural, Biological and Environmental Statistics* 27.2 (2022), pp. 364–381.
- [4] Keller, J. P., Zhou, T., **Kaplan, A.**, Anderson, G. B., and Zhou, W. "Tracking the transmission dynamics of COVID-19 with a time-varying coefficient state-space model". *Statistics in Medicine* 41.15 (2022), pp. 2745–2767.
- [5] Marchant, N.*, 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* 30.2 (2021), pp. 406–421.
- [6] 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.
- [7] **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.
- [8] **Kaplan, A.** and Hare, E. "Putting Down Roots: A Graphical Exploration of Community Attachment". *Computational Statistics* 34.4 (2019), pp. 1449–1464.
- [9] 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.
- [10] 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.

- [11] **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.
- [12] **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

[13] Taylor, I.*, **Kaplan, A.**, and Betancourt, B. "Fast Bayesian Record Linkage for Streaming Data Contexts" (2023+).

REFEREED CONFERENCE PRECEEDINGS

[14] 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

- [15] 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.
- [16] 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.
- [17] 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.

Grants & Contracts

As PI: Online Bayesian Learning for Streaming Entity Resolution (DoD via North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan and Brenda Betancourt (CSU: \$56,411). January 2022 - August 2022.

As PI: Streaming Record Linkage for Online Data Deduplication (DoD via North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan and Brenda Betancourt (\$130,352; CSU: \$90,624). January 2021 - December 2021.

As PI: Streaming Record Linkage for Online Data Deduplication (DoD via North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan and Brenda Betancourt (\$132,398; CSU: \$92,033). January 2020 - December 2020.

As PI: An Extensible Model for Deduplication of the GDELT Events Database (DoD via North Carolina State University Laboratory for Analytic Sciences). PI: Kaplan (\$23,996). August 2019 - December 2019.

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

Software

- [1] Taylor, I.*, **Kaplan, A.**, and Betancourt, B. bstrl: Bayesian Streaming Record Linkage. R package version 0.1.0. 2022.
- [2] Kaplan, A. and Bien, J. protoshiny: Interactive dendrograms for Visualizing Hierarchical Clusters with Prototypes. R package version 0.1.0. 2022.
- [3] Kaplan, A., Betancourt, B., and Steorts, R. C. representr: Create Representative Records After Entity Resolution. R package version 0.1.4. 2022.

- [4] Marchant, N., Steorts, R. C., and **Kaplan**, A. dblink: Distributed End-to-End Bayesian Entity Resolution. Scala package version 0.2.0. 2020.
- [5] **Kaplan, A.** and Nordman, D. conclique: Gibbs Sampling for Spatial Data and Concliques. R package version 0.1.0. 2017.
- [6] Hare, E. and **Kaplan**, **A.** intRo: Download and Run the intRo Statistical Software. R package version 0.1. 2017.
- [7] Kaplan, A. forestr: Random Forests with a User Created Splitting Criterion. R package version 0.0.0.9000. 2015.

INVITED TALKS

- [1] "Understanding Tree Demography Using Overlapping Lidar Scans and Spatial Entity Resolution". WNAR/IMS Annual Meeting. Invited Paper. WNAR. Anchorage, AK, June 2023 (Upcoming).
- [2] "Fast Bayesian Record Linkage for Streaming Data Contexts". 14th International Conference of the ERCIM WG on Computational and Methodological Statistics. Organized Invited Session. European Research Consortium for Informatics and Mathematics. London, UK, Dec. 2022.
- [3] "Fast Bayesian Record Linkage for Streaming Data". Record Linkage Interest Group of the ASA Seminar. Georgetown University, Massive Data Institute. Washington, D.C., USA (Remote via Zoom), July 2022.
- [4] "Exploring Interactive Dendrograms with Prototypes". *Invited Talk*. Iowa State University, Graphics Group. Ames, IA, USA (Remote via Zoom), Nov. 2020.
- [5] "Bayesian Canonicalization of Voter Registration Files". Symposium on Data Science and Statistics. Invited Session. ASA. Pittsburgh, PA, USA (Remote via Zoom), June 2020.
- [6] "Life after record linkage: Tackling the downstream task with error propagation". Bayes Comp. Invited Session. ISBA. Gainesville, FL, USA, Jan. 2020.
- [7] "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.
- [8] "Designing Modular Software: a Case Study in Introductory Statistics". *Joint Statistical Meetings*. Invited Session. ASA. Denver, CO, USA, July 2019.
- [9] "A Fast Sampler for Data Simulation from Markov Random Fields". Invited Session. The 28th Annual Conference of the International Environmetrics Society. Guanajuato, Gto., Mexico, July 2018.
- [10] "Model matters with restricted Boltzmann Machines". Keynote. The 1st Midwest Statistical Machine Learning Colloquium. Ames, IA, USA, May 2018.
- [11] "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.
- [12] "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] University of California Santa Cruz, Department of Statistics. Santa Cruz, CA, Mar. 2023 (Upcoming).
- [2] Dalhousie University, Department of Statistics. Halifax, NS, Canada, Mar. 2023 (Upcoming).
- [3] Colorado School of Public Health, Department of Biostatistics and Informatics. Denver, CO, Oct. 2022.
- [4] Colorado School of Mines, Department of Applied Mathematics and Statistics. Golden, CO, Apr. 2022.

- [5] McGill University, Department of Biostatistics. Montreal, QC, Canada (Remote via Zoom), Mar. 2022.
- [6] University of Iowa, Department of Statistics and Actuarial Science. Iowa City, IA, USA (Remote via Zoom), Sept. 2021.
- [7] University of Toronto, Department of Statistics. Toronto, ON, Canada (Remote via Zoom), Nov. 2020.
- [8] New York University, Center for Practice, Research at the Intersection of Information, Society, and Methodology. New York, NY, USA (Remote via Zoom), Oct. 2020.
- [9] The University of Wisconsin, Systems, Information, Learning and Optimization (SILO) Research Group. Madison, WI, USA, Nov. 2019.
- [10] Duke University, Department of Computer Science. Durham, NC, USA, Apr. 2019.
- [11] The University of California Irvine, Department of Statistics. Irvine, CA, USA, Feb. 2019.
- [12] The University of Kentucky, Department of Statistics. Lexington, KY, USA, Feb. 2019.
- [13] The Pennsylvania State University, Department of Statistics. State College, PA, USA, Feb. 2019.
- [14] Virginia Tech University, Department of Statistics. Blacksburg, VA, USA, Feb. 2019.
- [15] The University of Wisconsin, Department of Statistics. Madison, WI, USA, Feb. 2019.
- [16] Colorado State University, Department of Statistics. Fort Collins, CO, USA, Jan. 2019.
- [17] The University of Texas at Austin, Department of Statistics and Data Science. Austin, TX, USA, Jan. 2019.
- [18] Stanford University, Department of Statistics. Stanford, CA, USA, Jan. 2019.
- [19] Monash University, Department of Econometrics and Business Statistics. Melbourne, VIC, Australia, Jan. 2019.
- [20] Texas A&M University, Department of Statistics. College Station, TX, USA, Nov. 2018.
- [21] North Carolina State University, Department of Statistics. Raleigh, NC, USA, Oct. 2018.
- [22] Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2018.
- [23] North Carolina State University, Department of Statistics. Raleigh, NC, USA, Nov. 2017.
- [24] Centro de Investigación en Matemáticas. Guanajuato, Gto., Mexico, Oct. 2017.
- [25] Duke University, Department of Statistical Science. Durham, NC, USA, Apr. 2017.
- [26] Michigan State University, Department of Statistics and Probability. East Lansing, MI, USA, Feb. 2017.
- [27] University of Illinois, Urbana-Champaign, Department of Statistics. Chapaign, IL, USA, Feb. 2017.
- [28] University of Florida, Department of Statistics. Gainesville, FL, USA, Feb. 2017.
- [29] Rice University, Department of Statistics. Houston, TX, USA, Jan. 2017.
- [30] University of Virginia, Department of Statistics. Charlottesville, VA, USA, Jan. 2017.
- [31] Arizona State University, School of Mathematical and Statistical Sciences. Tempe, AZ, USA, Jan. 2017.
- [32] University of Massachussetts, Amherst, Department of Mathematics & Statistics. Amherst, MA, USA, Dec. 2016.
- [33] Iowa State University, Department of Statistics. Ames, IA, USA, Jan. 2015.

INVITED SHORT COURSES

Some of Record Linkage, Full-Day Workshop

 $Co\mbox{-}instructor$

US Census Bureau May 2018

Centro de Investigación de Matemáticas, A. C. February 2018

Machine Learning Day, Full-Day Event

Co-organizer

Duke University March 2018

D3 Workshop, Half-Day Workshop

Co-instructor

NORC at the University of Chicago October 2015

Week of R, Week-Long Workshop

Co-instructor

Iowa State University June 2013, 2014, 2015

Contributed Talks

- [1] "An improved sampler for recursive Bayesian inference". *Joint Statistical Meetings*. Topic Contributed Paper. ASA. Toronto, ON, Canada, Aug. 2023 (Upcoming).
- [2] "Counting Casualties in the Syrian Civil War with Bayesian Record Linkage". Joint Statistical Meetings. Topic Contributed Paper. ASA. Vancouver, BC, Canada, July 2018.
- [3] "Population Sized Graphical Record Linkage". *ISBA World Meeting*. Contributed Member Paper. ISBA. Edinburgh, UK, June 2018.
- [4] "A Simple, Fast Sampler for Simulating Spatial Data and Other Markovian Data Structures". Joint Statistical Meetings. Contributed Paper. ASA. Baltimore, MD, USA, Aug. 2017.
- [5] "An exposition on the propriety of restricted Boltzmann machines". *Joint Statistical Meetings*. Contributed Paper. ASA. Chicago, IL, USA, July 2016.
- [6] "Introducing Statistics with intRo". The R User Conference. Contributed Paper. Stanford, CA, USA, June 2016.
- [7] "Introducing Statistics with intRo". Joint Statistical Meetings. Topic Contributed Paper. ASA. Seattle, WA, USA, Aug. 2015.
- [8] "gravicom A web-based tool for community detection in networks". *Joint Statistical Meetings*. Contributed Paper. ASA. Boston, MA, USA, Aug. 2014.

Contributed Posters

- [1] "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] "Population Sized Record Linkage". New Researchers Conference. Contributed Poster. Institute of Mathematical Statistics. Vancouver, BC, Canada, July 2018.
- [3] "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] "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] "Visualizing Linked Data Sources for the National Children's Study". Conference on Statistical Practice. Contributed Poster. ASA. San Diego, CA, USA, Feb. 2016.

- [6] "Mathematical Self-Efficacy of Incoming Students at a Large Public University". Women in Statistics Conference. Student Poster. ASA. Cary, NC, USA, May 2014.
- [7] "Putting Down Roots: A Graphical Exploration of Community Attachment". *Joint Statistical Meetings*. Data Exposition Poster. ASA. Montreal, QC, Canada, Aug. 2013.
- [8] "Can you buy a president? Politics after the Tillman Act". *Joint Statistical Meetings*. Contributed Poster. ASA. Montreal, QC, Canada, Aug. 2013.

TEACHING EXPERIENCE

*Denotes courses developed.

CSU	STAA 577	Statistical Learning and Data Mining	S23
CSU	DSCI 445*	Statistical Machine Learning	F20, F21, F22
CSU	STAT 730	Advanced Theory of Statistics I	S20
CSU	STAT 400	Statistical Computing	F19, F20, F21, F22
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

Mentorship

Graduate Students

Cradado Stadonto	
Ian Taylor, PhD Advisor	Summer 2023 (Anticipated)
Lane Drew, PhD Advisor	Spring 2024 (Anticipated)
Connor Gibbs, PhD Committee Member,	Spring 2023 (Anticipated)
Justin Van Ee, PhD Committee Member	Spring 2023 (Anticipated)
Alex Fout, PhD Committee Member	Summer 2022
Xinyi (Lucy) Lu, PhD Committee Member	Summer 2021
Casey Schafer, GRA	Fall 2019

Undergraduate Students

Ryan Volkert (Colorado State University), Undergraduate Research		Spring 2021
Olivia Beck (Colorado State University), Honors Thesis Committee	Member	Fall 2019
Ritika Bharati (Duke University), Undergraduate Research	Spring 201	8 - Spring 2019
Srini Sunil (Duke University), Undergraduate Research	Fall 201	7 - 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

Departmental Service

Graduate Admissions Committee	Fall 2022 - Spring 2023
Undergraduate Committee	Spring 2022 - Spring 2023
Search Committee	Fall 2021 - Spring 2022
Search Committee	Fall 2020 - Spring 2021
Graduate Admissions Committee	Fall 2020 - Spring 2021
Data Science Committee	Fall 2019 - Spring 2021
Co-organize Statistical Learning and Data Science Journal Club	Fall 2020
Seminar Co-organizer	Fall 2019 - Spring 2020

 $University\ Service$

Data Science Research Institute Steering Committee

Fall 2020 - Present

Service to the Profession

Associate Editor

Technometrics January 2022 - Present

Reviewer

AIStats, American Political Science Review, ASA SBSS Student Paper Competition, Annals of Applied Statistics, Bayesian Analysis, Computational Statistics & Data Analysis, Environmental and Ecological Statistics, Journal of the American Statistical Association, Journal of Computational and Graphical Statistics, Journal of the Royal Statistical Society, Section A, Journal of Survey Statistics and Methodology, PLOS One, Statistics and Probability Letters, Statistical Science, Survey Methodology

NSF Panelist (DMS)

Conferences

Local Organizing Committee Member, The International Indian Statistical Association International Conference, Golden, CO, USA June 2023

Topic Contributed Session Organizer, Joint Statistical Meetings, Toronto, ON, Canada August 2023

Invited Session Organizer, WNAR, Anchorage, AK, USA

June 2023

Topic Contributed Session Organizer, Joint Statistical Meetings, Philadelphia, PA, USA (Session Cancelled due to COVID-19)

August 2020

Co-organizer, Symposium for Data Science & Statistics, Pittsburgh, PA, USA

Invited Session Chair, Joint Statistical Meetings, Denver, CO, USA

August 2019

Contributed Session Organizer, Joint Statistical Meetings, Vancouver, BC, Canada August 2018

Statistics in the Community (StatCom)

Network Outreach Coordinator

2015 - 2017

Executive Committee, ISU Chapte	Executive	Committee,	ISU	Chapte
---------------------------------	-----------	------------	-----	--------

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

Computing

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