

# Jason (Zhihang) Dong

Ph.D. Student | Looking for Research/Data/Machine Learning Scientist Intern Opportunities

in [linkedin.com/in/zhihangdong](https://www.linkedin.com/in/zhihangdong) | [github.com/zdong1](https://github.com/zdong1) | +1 814 548 6383 | [zdong@uw.edu](mailto:zdong@uw.edu)

📍 211 Savery Hall, Box 353340, Seattle, Washington, USA 98115 | <https://zdong.org>

## 🌐 Research Interests

Statistical applications of **AI and machine learning** as they apply to **social science questions**, specifically **online & adaptive learning, graphical models, geometric data analysis** (e.g. manifold learning), **Natural Language Processing** (QA and Information Extraction). Additionally, I am interested in **statistical methodologies** (survey, spatial-temporal models) in **demography**.

## ☰ Skills

**Note :** Skills with **bold** mean 50+ hours experience (400+ for languages) & ranked with familiarity

Computer Languages : **Python, R, Java**, Scala, C(Statistics Libraries)

Machine Learning Toolkits : **Spark, Torch, xgBoost, caffé, Tensorflow, H2O**, Hadoop, CTNK, EMR, Keras

Databases : **Neo4j, Cassandra, MongoDB, postgre, mySQL, RDS, mySQL, Hive**, CouchDB, Lucene

Natural Language Processing : **Tika, NTLK, SpaCy, TextBlob**, Stanford Core NLP

Softwares : **ArcGIS, Stata, tableau**, GeoDa, SAS

## 📁 Projects

Current  
June, 2017

Swiss Cell Data Record Project | Machine Learning, Mobility, GPS

- > Managed 100G+ flow of cell phone GPS record through AWS and cluster service
- > Wrangled the data to overcome its unstructured, noisy and space-time properties
- > Wrote 3000+ LOC to implement neural network and topological data methods
- > Modeled activity space, mobility and social interactions with innovated convergence algorithm

[R](#) [AWS](#) [d3](#) [Tensorflow](#) [TDN](#) [Python](#) [SQL](#) [Demography](#) [Mapping](#)

Current  
January 2018

Knowledge Transfer and Predictive QA via Reinforcement Learning and GANs | NLP, ML, Deep Learning

- > Predicted and Completed paths in Knowledge Graph with Reinforcement Learning, Generative Adversarial Networks (GAN) and Multi-layer bidirectional long short-term memory network (LSTM)
- > Designed end-to-end training of multiple documents across the Document Retriever and Document Reader pipeline for Question & Answer (QA) NLP Tasks
- > Reduced the Multisense Word Embedding problems by a word-swap validation procedure

[GAN](#) [Reinforcement Learning](#) [SpaCy](#) [Python](#) [NLP](#) [NTLK](#) [caffé](#) [LSTM](#)

May 2017  
January 2017

Neighborhood Sensing : Investigating Crimes with Network Effects | Space-Time Model, Mapping, API

- > Scrapped and Mapped Police data into City Gov APIs
- > Created Integrative **R** applications generate dynamic statistical models to map the measured network effects on neighborhood safety

[R](#) [Neighborhood Effects](#) [Time Series](#) [Social Network](#)

Mar 2018  
May 2017

Anomaly Detections of NYC Taxi Ridership using Adapted Bayesian Neural Network | ML, Methodology,

- > Designed an adapted Bayesian neural network with benchmarked SARIMA time series model
- > Improved predicted accuracy by 4% and implemented the inference system on Python
- > Applied manifold learning methods to conduct dimensionality reduction

[Bayes](#) [Time Series](#) [SARIMA](#) [Anomalies](#) [Python](#) [LSTM](#)

## 📁 Work Experience

September, 2017  
June, 2017

Center for Studies in Demography and Ecology, Summer Research Assistant, Seattle, WA

- > Provided 400+ hours of statistical/programming supports to five different mini-projects; Managed data from multiple clients
- > Implemented OD Matrix Mapping in GIS; Wrote Python programs to identify fake patient ID in the system with < 0.1% error rate
- > Wrote 800+ LOC; Worked with **leaflet.js** on interactive mapping

[R](#) [AWS](#) [d3](#) [leaflet](#) [Python](#) [SQL](#) [Demography](#) [Mapping](#)

Current Summer 2015	<b>Teaching Assistant and Research Assistant, PSU &amp; UW, Various Places</b> <ul style="list-style-type: none"> <li>&gt; Assisted and co-taught undergraduate-level methodological and substantive courses; Shared responsibilities for lectures, exams, assignments; Best Rating : <b>4.4/5.0</b>, 15 Summer; Refs Available</li> <li>&gt; Helped design experiments in Media Effects Lab at PSU; Conducted literature review and data analyses in <b>multiple disciplines</b></li> </ul> <div style="border: 1px solid gray; padding: 2px; display: inline-block;">Experimental Design</div>
May 2015 August 2015	<b>Undergraduate Summer Data Analyst, Survey Research Center of SSRI, University Park, PA</b> <ul style="list-style-type: none"> <li>&gt; Missing Data Info Retrieval for 100,000+ survey responses</li> <li>&gt; Implemented algorithms to detect false survey responses</li> </ul> <div style="display: flex; gap: 5px;"> <div style="border: 1px solid gray; padding: 2px; display: inline-block;">SAS</div> <div style="border: 1px solid gray; padding: 2px; display: inline-block;">SPSS</div> <div style="border: 1px solid gray; padding: 2px; display: inline-block;">Stata</div> </div>

## Languages

English	● ● ● ● ●
Chinese	● ● ● ● ●
Japanese	● ● ○ ○ ○

## + Honors & Awards

- > **Clarence and Elissa M. Schrag Endowed Fellow**  
University of Washington, 2016 (2 per Ph.D. Cohort)
- > **UW Cloud Computing Credits Research Awards**  
University of Washington, 2016
- > **Undergraduate Research Travel Awards**  
Penn State University, 2014 - 2016

## Education

2020/12 (est.)	Ph.D. — University of Washington, Seattle (CSSS Option, Sociology) <b>Areas</b> : Machine Learning, Social Sensing & Geometric Data Analysis <b>Advisor</b> : Prof. Adrian Dobra, Prof. Ross Matsueda (co-chair)
2019/3 (est.)	M.S. Statistics — University of Washington, Seattle
2016/05	B.A. — Pennsylvania State University, Sociology with Concurrent Majors and Minors in Statistics, Geography, Media Studies, Psychology (5 majors and 3 minors)

## “ Publications

### — — Conferences — —

- > **Dong Zhihang**, Yen-Chi Chen and Adrian Dobra (2018). “Projecting the Short-term Population Mobility using Cell Data Records”. *Accepted to 2018 Joint Statistical Meetings (JSM)*.
- > **Dong Zhihang** and Tongshuang Wu (2018). “Benchmarking Open Source NoSQL Databases Performance on NLP Queries”. *Work in Progress*.
- > **Dong Zhihang** (2017). “Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling”. In : *International Conference on Population Geography*.
- > — (2016). “Theorizing Urban Neighborhoods : Mapping the Interneighborhood and Intranighborhood Networks and Criminogenic Factors on Street Crime Victimization”. In : *American Society of Criminology Annual Meeting*.

### — — Journals — —

- > Dong, Zhihang. “Modeling Age Homogeneity : Age Homogamy And Marital Happiness Over The Life Course”. *Undergraduate Thesis, Under Review by JMF*.

## Course Works

- > 0. STAT 535 : **Statistical Learning** [Link]
- > 1. CSE 599i : **Online and Adaptive Learning** [Link]
- > 2. STAT 564 : **Bayesian Statistics**[Link]
- > 3. MATH 515 : **Optimization**[Link]
- > 4. CSE 547 : **Machine Learning for Big Data**[Link]
- > 5. EE 576 : **Computer Vision**[Link]
- > 6. CSE 599d : **Advanced NLP Methods** [Link]
- > 7. CSE 544 : **Database Management** [Link]
- > 8. CSE 512 : **Data Visualization** [Link]
- > 9. CSE 599d1 : **Advanced Topics NLP** [Link]
- > 10. EE 595 : **Representation Learning** [Link]
- > 11. LING 575 : **Novel NLP Applications** [Link]

## + Labs & Working Groups

- > **Geometric Data Analysis Reading Group**  
2018 – Current [Link]
- > **Working Group for Applied, Bayesian and Computational Statistics**  
2016 – Current [Link]
- > **Context Working Group (Sociology)**  
2016 – Current POC : Prof. Kyle Crowder
- > **Human Factors in GI Science Lab, PSU Geography**  
2014 – 2015 [Link]