

# JASON (ZHIHANG) DONG

Ph.D. Student & Seeking Data/Applied/Research/ML Scientist Internship

@ zdong@uw.edu    +1-814-548-6383    211 Savery Hall, Box 353340, Univ. of Washington    Seattle, WA, USA    zdong.org  
@zdong6    linkedin.com/in/zhihangdong    github.com/zdong1

## EXPERIENCE

### Swiss MDC Cell Data Records Project

Collaborated Research Project [Link](#)

March 2017 – Ongoing    Seattle, WA

- Managed and warehoused 100G+ flow of data through cloud computing services and built new big data database
- Wrangled, Cleaned and Engineered the data to overcome its unstructured, noisy and space-time properties
- Wrote 3000+ LOC to implement statistical and machine learning (kernel, tensor and other) methods
- Modeled routine activity space and human mobility and social interactions with consultations on current literature

### Summer RA on Spatial Methodology

Center for Studies in Demography and Ecology [Link](#)

May 2017 - Sep 2017    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

### Undergraduate Summer Data Analyst

Survey Research Center affiliated to Penn State SSRI

May 2015 - August 2015    University Park, PA

- Missing Data Info Retrieval for 100,000+ survey responses
- Implemented algorithms to detect false survey responses

### Teaching Assistant and Research Assistant

Penn State & University of Washington

May 2015 - Ongoing    Multiple Cities

- Assisted and co-taught undergraduate-level methodological and substantive courses; Shared responsibilities for lectures, exams, assignments; Best Rating: 4.4/5.0, 15 Summer; Refs Available
- Helped design experiments in Media Effects Lab at PSU; Conducted literature review and data analyses in multiple disciplines

## SKILLS/ TOOLKITS

Machine Learning: Spark Hadoop CNTK Hive Torch  
Tensorflow MXNet H<sup>2</sup>O Keras xgB caffe Lucene  
NLP: Tika NTLK SpaCy TextBlob Stanford Core NLP  
Softwares: ArcGIS LaTeX Excel GeoDa tableau SAS d3  
Databases: mySQL noSQL Mongo Cassandra Pig RDS

## EDUCATION

Ph.D., Research Areas: Social Statistics & Machine Learning

University of Washington, Seattle

July 2016 – June 2020 (Expected)

Advisor: Adrian Dobra, Ross Matsueda

B.A. Sociology + 4\*\*

Penn State University, University Park

August 2012 – May 2016

\*\*with concurrent majors/minors in Statistics/Geography/Psychology/Media Effects

## RESEARCH INTERESTS

### Machine Learning

Online & Adaptive Learning Models, Topological Data Analysis (Manifold Learning), Deep Learning

### Statistical Methodology

Space-Time Modeling, Graphical Models

### Social Sensing

Neighborhood Effects, Social Network

### Natural Language Processing

Semantic Analysis, Question Answering

## COMPUTER LANGUAGES

R, Python, SQL    ●●●●●●  
Java, Scala    ●●●●●●  
C(Stats Lib), JavaScript    ●●●●●●

## COURSES

Machine Learning for Big Data  
CSE 547, Sp 16, UW [Link]

Online and Adaptive Learning  
CSE 599, Wi 18, UW [Link]

Bayesian Statistics  
STAT 564, Sp 17, UW [Link]

Optimization  
MATH 515, Au 17, UW [Link]

## PROJECTS

### False Survey Data Detection

**Individual Project** [Link \(By Request\)](#)

📅 December 2016 – Ongoing 📍 Seattle, WA

- A Bayesian neural network approach to detect false/inconsistent entries on survey responses using various techniques such as sentiment analysis, deep belief network (DBN)
- False rate lowered by 16% compared to the benchmark model

### Neighborhood Sensing Project

**Individual Project** [Link](#)

📅 December 2016 – Ongoing 📍 Seattle, WA

- Streamed and Mapped data into online models from City Gov APIs
- Created Integrative R applications generate dynamic statistical models to measure network effects on neighborhood safety

### Anomaly Detection in Time Series using Bayesian Neural Network with Benchmark ARIMA

**Class Project**

📅 January 2018 – Ongoing 📍 Seattle, WA

- Detect Anomalies of NYC Cab Ridership by incorporating seasonal ARIMA model into the Bayesian neural network approach

### "Demography Master" - A Knowledge Transfer Tool

**Capstone Project**

📅 January 2018 – Ongoing 📍 Seattle, WA

- Answer open-ended demographic-related questions from Wikipedia Corpus; Built a QA database tool using CouchDB
- Reduce the Multisense Embedding Problems by Reinforce Correct Answers to the Intent-by-Wording Frameworks

## PUBLICATIONS

### 📄 Journal Articles

- Dong, Zhihang. "Age Homogamy And Marital Happiness Over The Life Course: What is There To Explain?" In: *Under Review by JMF*.

### 👥 Conference

- Dong Zhihang, Yen-Chi Chen and Adrian Dobra (2018). "Projecting the Short-term Population Mobility using Cell Data Records". In: *Submitted to 2018 Joint Statistical Meetings (JSM)*.
- Dong, Zhihang (2017). "Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling". In: *ICPG, International Conference on Population Geography*.
- – (2016). "Theorizing Urban Neighborhoods: Mapping the Interneighborhood and Intranighborhood Networks and Criminogenic Factors on Street Crime Victimization". In: *ASC*.

## HONORS

### Clarence and Elissa M. Schrag Endowed Fellowship

**University of Washington**

📅 2016-17 📍 Seattle, WA

**Selectivity:** *Only 2 per Ph.D. Cohort*

### UW HCC Cloud Computing Credits Awards for Research

**University of Washington**

📅 October, 2017 📍 Seattle, WA

**Selectivity:** *Unknown*

### Undergraduate Research Travel Awards + Dean's List

**Pennsylvania State University**

📅 Multiple 📍 University Park

**Selectivity:** *Travel Awards: Unknown, Dean's List: High GPA*

## HUMAN LANGUAGES

English

Chinese

Japanese



## HOBBIES

### Music

Lead Drummer (Random Algorithm)

### Sports

Bowling, Archery, Hearthstone (Arena 12-win Achiever), Tennis, Badminton

## MORE COURSES

### Statistical Learning

STAT 535, Au 16, UW [Link]

### Databases

CSE 544, Wi 18, UW [Link]

### Statistical Demography

STAT 593D, Sp 17, UW [Link]

### Advanced Regression Methods for Independent Data

STAT 570, Au 17, UW [Link]

### Spatial Statistics

STAT 591A, Au 17, UW [Link]

### Big Data and Population Processes

SOC 538, Au 16, UW [Link]