Wangjing (Jing) Ke

PROFESSIONAL SUMMARY

- "Big data" analyst with extensive programming experience in R, SAS, Python, Java, Javascript, PHP, and MySQL
- In-depth hands-on knowledge of large-scale study design, data mining, and statistical modeling
- Quick learner with strong communication and teamwork skills

EDUCATION

University of Southern California
M.S. Biostatistics, **GPA: 3.86/4.0** — May 2016
Ph.D Student in Epidemiology — Aug 2013 - Dec 2015

University of Arizona
MPH Epidemiology, **GPA: 3.92/4.0** — 2013
B.S. Biology, **GPA: 3.90/4.0**, *Summa Cum Laude*, *Phi Beta Kappa* inducted — 2011

WORK EXPERIENCE

Department of Preventive Medicine, USC — Los Angeles, CA
Data Management Coordinator — Jan 2016 - present

Processed streaming data acquired through smartphones and wearable sensor devices.

- Independently developed Android app *madresGPS* to record AES encrypted geolocations in customizable intervals
- Built web interface using Javascript to:
  - connect to Google Firebase to access data collected from mobile app
  - calculate and track real-time participant compliance rate of randomly prompted surveys on their phones
  - track the participant status for follow-up appointments and data collection
- Developed pipelines using R, Java, SAS, to:
  - process real-time behavioral and biometric data on over 400 participants
  - build and administer database, and conduct preliminary data cleanning and manipulation for 40+ researchers and staffs
- Performed multilevel mixed effect analysis and built structural equation models on human health behaviors

Norris Cancer Center — Los Angeles, CA
Research Assistant, Project Leader — Aug 2013 – Dec 2015

Designed statistical analysis plans and used human genomic data to analyze heart disease risks

- Identified heart disease patients from medical records in various sources
  - Developed algorithms predicting heart disease propensity for 250,000 people based on their medical history
  - Built and validated predictive models based on a small sample and applied it to a large population
- Examined and evaluated various causes of heart disease in adults
  - Analyzed how multiple demographic and contextual variables (e.g. age, gender, smoking, and drinking) affect the likelihood of heart disease
  - Conducted multivariate regression, ANOVA, survival and longitudinal analysis to identify the high risk population
  - Used R and SAS for data management, A/B testing, and analyses
- Searched candidate genes across the entire human genome that caused heart disease
  - Found 15 genetic mutations that cause heart diseases in over 1 million people in the United States
  - Built and manipulated genomic data for over 20,000 people using MySQL
  - Estimated the risk of heart disease on 30 million DNA mutations using high performance computing
  - Developed machine learning algorithms to select disease causing genes
  - *Tools used:* SAS, R, Python, MySQL, Perl, and AWK

EXPERTISE

Programming: R, SAS, Stata, SPSS, Tableau, Matlab/Octave, Python, Java, Javascript, PHP, MySQL, LaTeX, Bash, GNU/Linux
Specialties: Big data analysis, machine learning, data mining, high performance computing, experimental design, predictive modeling, high-dimensional/spatial/longitudinal/survival/nonparametric analyses, linear/logistic regression, statistical analysis plans, statistical review & quality control, data quality review

PERSONAL

Hobbies: skiing, fishing, watching basketball, and making drinks (certified by the Tucson Bartending Academy)