

WANGJING KE

Los Angeles, CA

Tel: 520-360-8048 | Email: kwangjing@gmail.com | Website: <http://wangjingke.com/>
LinkedIn: <https://www.linkedin.com/in/wangjingke> | GitHub: <https://github.com/wangjingke>

PROFILE

Data scientist with hands-on experience in statistical analysis, mathematical modeling, machine learning, and database administration. Proficient R, Python, and Java programmer, providing more than 40 researchers with analytic insights, data management and presentation of statistical results for publication to aid in research activities.

WORK EXPERIENCE

University of Southern California

Los Angeles, CA

Data Scientist

2016 - Present

- Develop statistical and machine learning models, utilizing streaming data from wearable sensors, for human habit recognition and behavior intervention.
- Perform ANOVA, time series multivariate on human health behaviors to increase result accuracy.
- Create dashboards for data visualization to improve team efficiency by allowing users to calculate participant compliance rates, and update the participant status for follow-up appointments.
- Provide programming solutions to data collection and management, and statistical analysis, such as:
 - SQL databases for multiple research projects with strict attention to detail
 - *madresGPS*, an Android app in Java to record encrypted geolocations in customizable time intervals
 - R package to harmonize month-long time series behavioral, dietary and biometric measurements
 - Python module to process minute-by-minute physical activity data from accelerometers

Data Analyst

2013 - 2015

- Developed predictive model, analyzing disease propensity in 250,000 people using medical records.
- Performed logistic regressions and A/B testing to determine the variables that affect the likelihood of heart disease.
- Managed more than 20 TB whole genome sequencing data and conducted quality control procedures.
- Performed principal component analysis to reduce data dimensions, controlling for multiple correlation.
- Designed and built machine learning models, combining clustering algorithms and k-nearest neighbor (kNN) algorithms to scan 40 million DNA mutations, identifying more than 70 heart disease related genes.

University of Arizona

Tucson, AZ

Research/Teaching Assistant

2011 - 2013

EDUCATION

University of Southern California

Los Angeles, CA

Master of Science – Biostatistics (GPA: 3.86/4.0)

2016

Doctor of Philosophy (Candidate) – Epidemiology

2013 - 2015

University of Arizona

Tucson, AZ

Master of Public Health – Epidemiology (GPA: 3.92/4.0)

2013

Bachelor of Science – Biology (GPA: 3.9/4.0)

2011

- **Awards/Honors:** Graduated *Summa Cum Laude*, *Phi Beta Kappa* inducted

SKILLS

Programming: R, SAS, Python, Java, SQL, Hadoop, Tableau, Microsoft Office (Word, Excel, PowerPoint, Access, Outlook)

Statistics: A/B testing, predictive modeling, linear/logistic regression, time series, experiment design, data mining

Machine learning: classification, clustering, regression models, feature engineering, principal component analysis

Language: English, Mandarin Chinese