

# Sydeaka P. Watson, Ph.D.

[770-557-6364](tel:770-557-6364) | [sydeakawatson@gmail.com](mailto:sydeakawatson@gmail.com) | [Github Profile](#) | [Personal website](#)

---

## PROFILE

Data scientist with strong R programming skills seeking to connect with an innovative, data-driven company that leverages modern technology and machine learning to gain insight from data.

## TECHNICAL SKILLS

- Big Data Tools: H2o, Hadoop, Hive, RCloud, Spark (SparkR/Sparklyr), SQL (JDBC, Teradata, Aster)
- Programming/Version Control: GitHub, Python, R, SourceTree
- Data Analysis: Hypothesis testing, interval estimation, machine learning, regression
- Visualization: Shiny, R Markdown

## RECENT WORK EXPERIENCE

### Senior Data Scientist AT&T, Chief Data Office

Dec 2016 – Present

Collaborated with senior business leaders and technical team to identify opportunities for retail optimization; proposed methods for retail evaluation and extraction of actionable insights to drive revenue growth

- Designed pilot study for testing proposed solutions in retail locations
- Created visualizations to communicate study performance metrics to senior business leaders and technical team
- Performed data exploration and data cleaning to facilitate development of prediction models
- Developed and validated classification models using machine learning algorithms

### Research Associate (Assistant Professor) The University of Chicago, Biostatistics Laboratory

Aug 2011 – Nov 2016  
<http://bit.ly/spw-uchicago>

Consulted with over 110 biomedical research teams in The University of Chicago Medical Center, designing clinical research studies (sample size calculation and analysis plan development), analyzing study data, and summarizing results for the study team and greater scientific community.

- Developed a survival probability prediction equation for patients on the national lung transplant waiting list given missing covariate data and a large number of highly correlated features. Compared to other published survival prediction models for this population, survival estimates obtained using the novel approach more closely resembled Kaplan-Meier estimates at 1, 2, and 5 years.
- Developed a missing data imputation algorithm for interval censored count covariate data used to examine the relationship between overall survival and rate of disease progression.

### Graduate research assistant: Los Alamos National Laboratory Division of Theoretical Biology and Biophysics

May 2009 - Jul 2011  
<http://bit.ly/spw-hiv>

Biostatistician for an international multi-disciplinary, multi-institutional team of HIV vaccine researchers testing theoretical design strategies in a remote animal research laboratory. Duties include analysis of immune response data, presentation of study findings to research team, and production of high-quality figures for team leader's conference presentations and journal submissions.

## EDUCATION

- Ph.D. Statistics (May 2011), M.S. Statistics (Dec 2008): Baylor University, Waco, TX
- M.S. Mathematics (Dec 2003): Michigan State University, East Lansing, MI

\*\* Full listing of publications, conference presentations, teaching, and professional service available at [www.sydeaka.com](http://www.sydeaka.com).