+1 (510) 666 7852
awiedmer@awiedmer.net
artwr.github.io
linkedin.com/in/arthurwiedmer
www.github.com/artwr

Arthur Wiedmer

Education

2008–2014 **PhD in Environmental Engineering (Expected)**, *University of California*, Berkeley, CA
Minors : Statistics, Public Health.
Dissertation topic : Estimating radioactive contaminant inventory at the Savannah River Site F-area seepage basins

2007–2008 **MS in Environmental Engineering**, *University of California*, Berkeley, CA

2004–2008 **BS+MS, Engineering Science**, *Ecole Polytechnique*, Palaiseau, France
Emphasis : Populations Biology, Mathematics, Physics, Computer Science

Experience

2014–Present **Data Engineering Fellow**, *Insight Data Science*, Palo Alto, CA
- Project: Wikihub : Github analytics on Wikipedia Pages
  - Build a data pipeline and API to enable Wikipedia edit metadata (>200Gb) analytics.
  - Used Hive for aggregation over different time granularity, stored the results in Hbase.
  - Displayed results through an API built on top of Flask.

2008–2014 **Graduate Student Research Assistant**, *UC Berkeley(2008-2014)/Lawrence Berkeley National Lab(2009,2011-2013)*, Berkeley, CA
Dissertation related research :
- **Wrote ETL code in SQL and Python to ingest heterogeneous data sources** (Oracle DB, Access DB, flat files in various formats) into SQL server as part of the Advanced Science Capability for Environmental Management(http://esd.lbl.gov/research/projects/ascem/) project data management team
- Developed a schema for a MS SQL Server database with GIS capabilities, transforming the data for a PostgreSQL instance serving as a backend for browser visualization as part of a multidisciplinary team of 8 people
- Predicted residence time of contaminants at the site of study using spatial statistics
- Created visualizations (MATLAB, R) used as a support for conceptual understanding by research teams at Lawrence Berkeley National Lab

Side projects :
- **Designed and ran introductory tutorials to MongoDB (JS and PyMongo) and Hadoop** for a seminar organized by the Lawrence Berkeley Lab using sample datasets: (US Census 2010 for Hadoop and Reuters for MongoDB)
- Selected to present a poster on data ingest and QA/QC issues encountered during the ASCEM project for the Conference on Data Analysis in February 2012

Spring '10,'11,'13, Fall '13 **Graduate Student Instructor**, *University of California*, Berkeley, CA
- Teaching assistant for both undergraduate (Freshman Physics, Hydrology) and graduate (Environmental Physicochemical Processes) classes
- Managed discussion sections and lab sessions for around 50 students
- Monitored student progress and offered additional tutoring for students with difficulties

Skills

Recent Hadoop, Pig, Hive, Hbase, Cassandra, Storm, Kafka

Experienced R, Python, Scipy stack, Version Control, RDBMS (SQL Server, Postgres)

Used MATLAB/GNU Octave, MongoDB, Java

Languages English (Fluent), French (Native), German (Conversational)