

# Stephen Y. Hwang

(224) 730-9044  
stephenyounghwang@gmail.com  
gnawhnehpets.github.io  
linkedin.com/in/hwangstephen

## Skills

<b>Programming</b>	R (dplyr, tidyr, ggplot), Python, Perl, bash, Shiny, RMarkdown, knitr, Jupyter, HTML, CSS Working knowledge of Airflow, Spark, Hadoop
<b>Warehousing</b>	SQL, Snowflake; Working knowledge of Redshift, Hive, MongoDB, BaseX
<b>Other technologies</b>	Tableau, Jira, Confluence, Git, Unix/Linux environments, AWS, REDCap

## Professional Experience

**Senior Bioinformatics Scientist**, RTI International, Research Triangle Park, NC Jun 2015 - present

- Supervise multiple teams of data analysts as program manager and technical task lead on multiple projects
- Responsible for designing and building pipelines to extract, transform and load (ETL) data, enhance data interoperability, and automate warehousing solutions
- Use R to implement data transformations (dplyr, tidyr) and create novel visualizations (ggplot)
- Utilize Snowpipe to ingest Twitter JSON data into Snowflake, where data is queried and transformed
- Create intermediate tables from raw data to support ad-hoc analyses and reporting
- Design and create tools and pipelines to support analytics needs and facilitate analysis and insights
- Design and configure AWS framework to host ETL tools and Shiny applications
- Interface with both technical and non-technical stakeholders to resolve complex business and technical issues

**Data Engineer (Consultant)**, Animal Cancer DX, LLC, Raleigh, NC Jun 2021 - present

- Designed and created an ETL pipeline to aggregate and warehouse experimental data on AWS S3, and automatically generate and distribute reports with statistical analyses to stakeholders
- Designed and configured AWS framework (EC2, S3, RDS) to host a Shiny application
- Designed and built data browser dashboard using Shiny and ggplot

**Bioinformatics Data Scientist**, Johns Hopkins Medical Institutions, Baltimore, MD Aug 2013 - Jun 2015

- Completed 2-year data mining project in 10 weeks using R, SQL, and bash
- Designed and created an ETL batch process to stage public health datasets and ingest into a SQL database
- Used Python to implement data transformations (pandas) and create novel visualizations (matplotlib, ggplot)
- Designed and implemented analysis pipelines for large genomic datasets for exploratory analysis and publication
- Established unsupervised machine learning methods into quality control process

**Bioinformatics Analyst**, Johns Hopkins Medical Institutions, Baltimore, MD Oct 2011 - Aug 2013

- Designed and created an ETL pipeline to ingest and transform public health data sources for analytic use, increasing the efficiency by 700% over the legacy workflow

**Bioinformatics Researcher**, Putonti Lab, Loyola University, Chicago, IL Jan 2010 - Jan 2011

- Created pipeline to aggregate and analyze genomic data extracted from NCBI repository

## Education

**Johns Hopkins University** Baltimore, MD  
M.Sc. Bioinformatics Aug 2013

**Loyola University** Chicago, IL  
B.Sc. Bioinformatics May 2011