

# 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

<b>Senior Bioinformatics Scientist</b> , RTI International, Research Triangle Park, NC	Jun 2015 - present
<ul style="list-style-type: none"><li>• Supervise multiple teams of analysts and developers as program manager and technical task lead</li><li>• Responsible for designing and building pipelines to extract, transform and load (ETL) data, enhance data interoperability, and automate warehousing solutions</li><li>• Use R to implement data transformations (dplyr, tidyr) and create novel visualizations (ggplot)</li><li>• Utilize Snowpipe to ingest Twitter JSON data into Snowflake, where data is queried and transformed</li><li>• Create intermediate tables from raw data to support ad-hoc analyses and reporting</li><li>• Design and create tools and pipelines to support analytics needs and facilitate analysis and insights</li><li>• Design and configure AWS framework to host ETL tools and Shiny applications</li><li>• Interface with both technical and non-technical stakeholders to resolve complex business and technical issues</li></ul>	
<b>Data Engineer (Consultant)</b> , Animal Cancer DX, LLC, Raleigh, NC	Jun 2021 - present
<ul style="list-style-type: none"><li>• 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</li><li>• Designed and configured AWS framework (EC2, S3, RDS) to host a Shiny application</li><li>• Designed and built data browser dashboard using Shiny and ggplot</li></ul>	
<b>Bioinformatics Data Scientist</b> , Johns Hopkins Medical Institutions, Baltimore, MD	Aug 2013 - Jun 2015
<ul style="list-style-type: none"><li>• Completed 2-year data mining project in 10 weeks using R, SQL, and bash</li><li>• Designed and created an ETL batch process to stage public health datasets and ingest into a SQL database</li><li>• Used Python to implement data transformations (pandas) and create novel visualizations (matplotlib, ggplot)</li><li>• Designed and implemented analysis pipelines for large genomic datasets for exploratory analysis and publication</li><li>• Established unsupervised machine learning methods into quality control process</li></ul>	
<b>Bioinformatics Analyst</b> , Johns Hopkins Medical Institutions, Baltimore, MD	Oct 2011 - Aug 2013
<ul style="list-style-type: none"><li>• 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</li></ul>	
<b>Bioinformatics Researcher</b> , Putonti Lab, Loyola University, Chicago, IL	Jan 2010 - Jan 2011
<ul style="list-style-type: none"><li>• Created pipeline to aggregate and analyze genomic data extracted from NCBI repository</li></ul>	

## Education

<b>Johns Hopkins University</b> M.Sc. Bioinformatics	Baltimore, MD Aug 2013
<b>Loyola University</b> B.Sc. Bioinformatics	Chicago, IL May 2011