

# David Nicholson

dnicholson329@gmail.com • 412-607-6313 • 11291 Chatterly Loop Apt 104, Manassas VA 20109

## Summary

Data Scientist with 13+ years of programming experience and 5+ years of experience in data analytics and visualization. Most of my data analytics experience has utilized language models and document embeddings to gain further insight into biomedical research.

## Skills & Proficiency

Github • Python • R • SQL • Google BigQuery • Data Analysis • Machine Learning • Deep Learning • Natural Language Processing • Transformers • Large Language Models • Text Mining • Topic Modeling • Clustering • Knowledge Graphs • Document Embeddings • Data Visualization • ETL Pipelines • API Framework • Databases • Algorithms • Software Development • Parallel Processing • Google Cloud Platform • Dashboards • Continuous Integration (CI/CD) • Docker

## Professional Experience

### **Data Scientist**

#### **Digital Science & Research Solutions, Ltd.**

June 2022 - Present

- Generated a software package make using a knowledge graph a lot easier.
- Constructed a pipeline that clustering and categorized documents for clients affiliated with the NIH
- Maintained an ETL Pipeline for NIH
- Constructed a textual processing pipeline to detect cancer drug treatments using a 64k biomedical document set for pharmaceutical clients.
- Backtested vector databases to determine which database is most optimal in dealing with 10M+ document embedding vectors.
- Constructed a pipeline that used a deep learning and dimensionality reduction models to uncover research topics and trends within a 20M+ document set for government clients.

### **Graduate Research Scientist**

#### **University of Pennsylvania**

August 2016 - June 2022

- Designed and implemented parallel processing pipelines that achieved a 3x speed-up when analyzing terabytes of biomedical text.
- Used weak supervision for a 1.5x speed-up when training deep learning models (recurrent neural networks and transformers) to extract biomedical relationships from biomedical text.

- Applied a k-nearest-neighbor model to provide scientists with a web service that identifies a listing of journals linguistically similar to a preprint of interest.
- Applied a time series analysis to discover over 20,000 different timepoints where words have changed their semantic meaning.

## Publications

- **Unmasking The Language Of Science Through Textual Analyses On Biomedical Preprints And Published Papers**  
**Nicholson, D. N.** (2022)
- **Changing Word Meanings in Biomedical Literature Reveal Pandemics and New Technologies**  
**Nicholson, D. N.** Alquaddoomi, F., Rubinetti, V., Greene, C. S. (2023)
- **Characterization of the Genome and Silk-gland Transcriptomes of Darwin's Bark Spider (*Caerostris darwini*)**  
Babb, P. L., Gregorič, M., Lahens, N. F., **Nicholson, D. N.**, Hayashi, C. Y., Higgins, L., Kuntner, M., Agnarsson, I., Voight, B. F. (2022)
- **Examining Linguistic Shifts between Preprints and Publications**  
**Nicholson, D. N.**, Rubinetti, V., Hu, D., Thielk, M., Hunter, L. E., Greene, C. S. (2022)
- **Expanding a Database-derived Biomedical Knowledge Graph via Multi-Relation Extraction from Biomedical Abstracts**  
**Nicholson, D. N.**, Himmelstein, D. S., Greene, C. S. (2022)
- **Ten important roles for academic leaders to promote equity, diversity, and inclusion in data science.**  
Moore JH, Truong VQ, Robbins AB, **Nicholson D. N.**, Williams-Devane CL. (2021)
- **Constructing Knowledge Graphs and Their Biomedical Applications**  
**Nicholson, D. N.**, Greene, C. S. (2020)
- **The Nephila Clavipes Genome Highlights the Diversity of Spider Silk Genes and their Complex Expression**  
Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., **Nicholson, D. N.**, Kim, E.J., Hogenesch, J.B., Kuntner, M., Higgins, L., Hayashi, C. Y., Agnarsson, I., Voight, B.F. (2017)

## Education

**Doctor of Philosophy (Ph.D.), Genomics and Computational Biology;** University of Pennsylvania (Philadelphia, PA)

**Postbaccalaureate Program (Penn Prep);** University of Pennsylvania (Philadelphia, PA)

**Bachelor of Science, Computer Science;** University of Maryland Baltimore County (Baltimore, MD)