# Y. Shirley Li

765-430-8920 | li1050@purdue.edu | LinkedIn | GitHub

## Summary

GIS Specialist and Ph.D. candidate with 9+ years in analytics and statistical modeling, skilled in Python, R, and SQL. Applying ML, data science, and operations research to optimize systems and deliver actionable insights.

## Experience

**Purdue University Libraries and School of Information Studies** — West Lafayette, IN
*GIS Specialist/Instructor/Analyst* — *2016 - 2025*

- Partnered with senior faculty and research leadership to design and deploy data pipelines and scalable geospatial infrastructure, enabling 20+ cross-disciplinary projects
- Developed and deployed interactive dashboards and custom web apps to support geovisualization across campus
- Led workshops and certificate programs, driving adoption of data science and geospatial AI across disciplines

## Education

**Purdue University** — West Lafayette, IN
*Ph.D. in Sustainability Engineering and Environmental Engineering* — *2020 – 2026*
*Graduate Certificate in Applied Statistics* — *2018 – 2020*
*M.S. in Civil Engineering* — *2012 – 2014*

**University of Waterloo** — Waterloo, ON
*Bachelor's in Geomatics, Minor in Computer Science* — *2010 – 2012*

**Wuhan University** — Wuhan, China
*Bachelor's in GIS (dual degree)* — *2008 – 2010*

## Projects

**Demand Prediction for Bike Share System Expansion** | *ML pipeline, XGBoost, Docker, AWS, Leaflet*

- Developed ML pipeline to predict station demand using XGBoost/linear regression with 4+ years of trip data
- Engineered spatial features from census, POI, and infrastructure layers, and optimized performance
- Deployed a containerized model as a serverless FastAPI inference service on AWS (ECR, Lambda, API Gateway)
- Built an interactive web map (Leaflet hosted on Amplify) for users to receive real-time demand predictions

**Solar-Powered Bike Share Station Modeling** | *Python, Simulation, Optimization*

- Pioneered the first study to quantify overlooked battery replacement needs (published in *Applied Energy*, 2023)
- Developed a simulation framework in Python for PV-battery charging, discharging, and replacement cycles
- Recommended optimization strategies that improved energy independence by 30% and reduced downtime by 50%

**Measuring and Modeling Equity in Bike Share Systems** | *Metric design, Regression model*

- Designed novel equity performance metrics to identify and quantify service gaps in bike share systems
- Analyzed millions of trip records and high-frequency station status data using regression models
- Estimated and assessed equity of rebalancing operations, revealing disparities in service distribution
- Provided actionable, evidence-based recommendations for integrating performance metrics into equity policies

**Alien Forest Pest Explorer** | *SQL, JavaScript, Dashboard, Stakeholder collaboration*

- Developed the first nation-wide, county-scale interactive dashboards, enabling users to track spread and impacts
- Built data workflows to manage and visualize invasive species datasets for effective forest management and policy
- Engaged stakeholders to ensure the platform delivers actionable, accessible insights for diverse user needs

## Technical Skills

**Languages**: Python, SQL, R, JavaScript, HTML/CSS, MATLAB
**Machine Learning & Data Science**: XGBoost, scikit-learn, PyTorch, TensorFlow, regression models, statistical modeling, hypothesis testing, A/B testing, causal inference, metric design, feature engineering, model evaluation
**Data Engineering & MLOps**: ETL, Docker, FastAPI, Git/GitHub, CI/CD (GitHub Actions), AWS (Lambda, API Gateway, ECR, S3), Google Cloud Platform, Spark (PySpark), Sedona, SQL Server, database management
**Visualization & Web Apps**: Tableau, Power BI, Matplotlib, interactive dashboard design, data storytelling, Leaflet
**Geospatial & Remote Sensing**: GeoPandas, ArcPy, ArcGIS Pro/Enterprise/Online, QGIS, Google Earth Engine