VANSH GUPTA

2 Leighton Street, Apt 613 ◇ Cambridge, MA 02141
(917) 615 9773 ◇ vg224@cornell.edu ◇ vansh-gupta.com
Innovation ◇ Urban ◇ Development ◇ Strategy ◇ Education ◇ Healthcare IO

EDUCATION

Cornell University

B.S. in Biometry and Statistics 2023
Ann S. Bowers College of Computing and Information Science **GPA:** 3.95
Concentrations: Mathematical Statistics; Computational Statistics & Data Management

B.S. in Applied Economics and Management 2023
Charles H. Dyson School of Applied Economics and Management **GPA:** 3.92
Research Honors (Distinction) in Social Science
Dyson Diversity Scholar

RELEVANT COURSEWORK

Graduate Mathematics: Real Analysis (A), Honors Introduction to Analysis (A+), Stochastic Processes
(Theory) Stats: Econometrics, Theory of Statistics, Advanced Biological Stats, Probability & Inference
(Applied) Stats: Stat. Computing, Risk Sim. & Monte Carlo, MATLAB Computing, Analytic Modelling
Computer Science: Object-Oriented Programming & Data Structures, Database Management
Fundamental Math & Economics: Linear Algebra, Multivariable Calculus, Intl. Trade & Finance, Intermediate Microeconomics, Intro Microeconomics, Intro Macroeconomics, Thinking With Mathematical Perspectives
Strategy, Business, & Law: Strategy, Finance, Business Law, Managerial Economics

RESEARCH PROJECTS

- **Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsors Program**
  NBER Working Paper 31668 with co-authors Annamaria Conti, Jorge Guzman, and Maria Roche
  - Built data pipelines leveraging GitHub REST & GraphQL APIs to form a panel dataset of user activity.
  - Created & launched a survey aimed at understanding incentives behind users' contributions to open source.
  - Conducted power analyses and preliminary regressions to estimate changes in innovation post-sponsorship.
- **State Capacity & Responses to Natural Disasters: Evidence from Brazilian Municipalities**
  Work in progress with co-authors Michael Best, Renata Lemos, and Daniela Scur
  - Coordinating the web-scraping of ~25 million PDF files in Portuguese and their processing.
  - Helping implement machine learning and clustering algorithms, managing and tracking big data, natural language processing, and the use of LLMs to speed up data processing and reduce costs.
  - Produced geospatial data visualizations in Python & Tableau to create interactive graphics, and helping conduct a discrete choice experiment and exploratory analyses.
- **Approaches and Resources for Improved Student Outcomes: Evidence from Brazilian Schools**
  Independent Working Paper
  - Ideated the research from scratch, & submitted as my honors thesis to Department of Global Development.
  - Conducted rigorous analyses to understand the effect of pre-existing resources on the efficacy of different strategies meant to improve student outcomes, using the Strategic Treatment Effects framework as a base.
  - Utilized an open source statistical analysis R package (STE) that I helped build to estimate treatment effects of strategies, which leverages random forest generated propensity scores in the process.

RESEARCH EXPERIENCE

- **Antitrust & Competition Economics - Associate & Data Scientist, Charles River Associates**
  June 2023 - Present; Worked 2400+ Hours Annually On Rigorous Economic Research
  - Produce economic analyses evaluating the competitive effects of healthcare mergers in the US, building on research from industrial economics, urban economics, and healthcare economics.
  - Author white-papers presented to the U.S. Department of Justice and juries to discuss the efficacy of legal frameworks that govern antitrust enforcement policies in the US for the healthcare industry.
  - Hold a dual role as a data scientist/software engineer as well, creating interactive EXE tools to enable economic analyses with large-scale data with ease - including software enabling robust geo-spatial analyses.
  - Assist with the adoption of cloud computing practices for economic analyses involving big data.
- **Dean's Summer Research Fellow, Columbia Business School**
  March 2022 - July 2022; Volunteered July 2022 Until Present
  Worked under the guidance of Jorge Guzman, co-authoring with Maria Roche and Annamaria Conti for Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsors Program.
  - Created data pipelines to gather data at a large scale and authored an open-source R package (STE).
  - Contributed to, and presented on projects in Strategy, Entrepreneurship, and Digital Innovation.
- **Research Assistant, Cornell Dyson School of Applied Economics & Management**
  Feb 2021 - May 2023; Volunteered July 2024 Until Present
  Worked under the guidance of Daniela Scur, co-authoring with Michael Best and Renata Lemos for State Capacity & Responses to Natural Disasters: Evidence from Brazilian Municipalities.
  - Spearheaded a large-scale data collection & analysis effort to digitize millions of PDFs from Brazil.
  - Worked briefly with Benjamin Leyden to study competition & innovation, with a focus on digital industries.
- **Data Reproducibility Researcher, American Economic Association**
  May 2021 - August 2021
  Responsible for reproducing results in manuscripts submitted for review to AEA journals under the AEA Data Editor, Lars Vilhuber.
  - Debugged and reproduced code in 30+ manuscripts, and verified the validity of the data.
  - Generated analysis replication reports, which were then used to make review decisions for the manuscripts.
  - Authored an open-source data utility tool which is used by AEA to upload reproduction files to Zenodo.
- **NLP Research Assistant, Cornell University**
  Jan 2021 - October 2021
  Worked under the guidance of Reza Moghimi, studying Twitter sentiment's effects on the financial markets.
  - Conducted literature reviews for the paper, studying NLP algorithms for sentiment analyses.
  - Helped in the development of Python code to collect the data and conduct the analysis.

TALKS

Theoretical and Applied Development - Presentations On Learning & Experiences
Cornell University, Department of Global Development, Ithaca, NY (April 20, 2023).

Development, Firms, & Labor Research Working Group Presentation
Cornell University, Dyson School of Applied Economics & Management, Ithaca, NY (March 17, 2023).

Strategy & Innovation Research Presentation
Columbia University, Columbia Business School, New York, NY (July 28, 2022).

TEACHING

TA - CS 2110: Object-Oriented Programming & Data Structures *(Feb 2022 - May 2022)*

TA - CS 1112: Introduction to Computing Using MATLAB *(Feb 2021 - Dec 2021)*

SOFTWARE DEVELOPMENT

- **[Open Source] R Package: STE**
  R Package to implement the framework developed in Treatment Effects in Managerial Strategies, using machine learning (random forest) in a causal inference setup to estimate strategic treatment effects.
- **[Open Source] Python: Upload-To-Zenodo**
  Python code to parse directories and upload the replication code to Zenodo. This is used by the American Economic Association replication team for smoother synchronization of their files.
- **Python & STATA Package, EXE: Google Maps API - Geocoding & Distance Matrix Calculator**
  Used Python, Azure, and Google Cloud to create and maintain a secure environment that authenticates a user within the organization and provides an easy interface for geocoding and routing calculations for coordinates.
- **Python Package: Divestment Optimization Calculator**
  Developed a tool leveraging algorithms to compute large-scale optimization solutions & recommend divestitures for multi-market mergers, particularly those relying on geographic market definitions (such as hospitals).
- **Python Package, EXE: Interactive Map Creator**
  Pioneered the use of fully customizable and native interactive maps as legal support exhibits. Innovated internal tools to make this technology readily available to all employees within the firm without prior training.

PROFESSIONAL WORK EXPERIENCE

- **Ascenta Management Consulting**
  Founder & CEO, August 2019 - May 2023
  - Recognized as one of India's Top 10 Consultants (CEOInsights Magazine).
  - Led teams of volunteer consultants working pro-bono with Indian Small and Medium Enterprises to improve their operations workflows, cut costs, expand their clientele, & provide effective data-driven strategies.
  - Consulted 20+ businesses using data to aid resource mobilization & expand revenue streams for the firms.
- **Cornell Business Analytics**
  Vice President, August 2020 - May 2023
  - Motivated team-members to ensure the nurturing of innovative ideas and the timely delivery of deliverables.
  - Led teams of analysts organizing the recruitment screen of 180+ applicants, designing innovative challenges.
  - Optimized inventory management for a client using operations research theory and Python.
  - Leveraged data to create visual insights & Tableau dashboards for a startup to find hotspots for its launch.
- **Meta**
  Big Data & Market Intelligence Analyst Extern, February 2022 - May 2022
  - Collaborated with teams to use data-driven insights for suggesting clients to Meta's Global Business Group.
  - Built machine learning code (k-means clustering & random forests) for insights & predictive analytics.
  - Consulted internal teams at Meta, curating a client acquisition strategy for Small Business Clients Group.
- **Grant Thornton International LLP**
  Research Trainee – Global Research Centre, December 2018 - January 2019
  - Developed a database of client information to streamline communication across Grant Thornton India by leveraging Capital-IQ & Avention databases, implementing heavy use of SQL and VBA for analytics.
  - Analyzed companies to profile their strategies and risks, producing thorough SWOT analyses for industries.
  - Consulted with Australian clients and compiled market-competition research reports for them.

PROGRAMMING SKILLS

Scripting Languages: Python, Java, Visual Basic (VBA), JavaScript (Intermediate)
Statistical Analysis: STATA, MATLAB, R, SAS (Exposure), Fortran (Exposure)
Mapping & Spatial Data: ArcGIS Pro, Google Maps API, Folium/Leaflet (Interactive Maps)
Data Scraping: PDFPlumber, BeautifulSoup, Selenium
Big Data & Data Management: Microsoft Azure, Databricks, SQL, GraphQL, Excel, Google Cloud
Data Visualisation: Tableau, Adobe InDesign, kepler.gl

HONORS & MEDIA COVERAGE

- Distinction in Social Science research, with Magna Cum Laude honors at Cornell University.
- Recognized as one of India's Top 10 Business Consultants by CEOInsights Magazine India for my work with Ascenta Consulting.
- Research covered in the NBER Bulletin on Entrepreneurship.
- Research covered in the Harvard Business School Working Knowledge Newsletter.
- Dean's List for all my semesters at Cornell University.
- Recognized as a Dyson Diversity Scholar at Cornell University.

STANDARDIZED TESTING

- **Graduate Record Exam (GRE)** **Scores:** Verbal 165/170, Quant 170/170, Writing 6/6
  Date: Aug 2023
  Scored 165/170 (95th percentile) in Verbal skills, 170/170 in Quantitative skills, and 6/6 in Analytical Writing, which is limited to the top 1% of essay writers. A score of 335/340 is in the top 1% of scores.
- **Thinking Skills Assessment Test (TSA Oxford)** **Score:** 97th Percentile
  Date: Nov 2019
  Scored a 77.7 (97th percentile score) among all the applicants applying to the Universities of Oxford and Cambridge from all over the world on a standardized test. Achieved 96th percentile in "Problem Solving" and a 93rd percentile in "Critical Thinking".
- **SAT** **Score:** 1530/1600
  Date: Dec 2018
  Achieved an SAT score of 1530/1600 with a global percentile rank of 99th, an English (Evidence-Based Reading and Writing) score of 730/800, and a Mathematics score of 800/800.
- **TOEFL** **Score:** 109/120
  Date: Apr 2019
  TOEFL is the Test of English as a Foreign Language for non-native English speakers.

LANGUAGES & CITIZENSHIP

Languages: English (Native), Hindi (Native) **Citizenship:** India

INTERESTS

Transport Systems, Urban Design, Cartography, Aviation, Competition Law, Parkour, Hands-Free Biking