

# Leslie Huang

lesliehuang@nyu.edu

NYU Center for Data Science, 60 Fifth Avenue

leslie-huang.github.io • linkedin.com/in/huangleslie

---

## EDUCATION

**New York University** | PhD in Data Science, 2018– (*expected May 2021*)

- *Transferred PhD programs with full support for my interdisciplinary research.*
- Dissertation: “Contextual NLP tools for understanding political communication”
  - “Using BERT to characterize partisan connotation of words”
  - “A semantic and stylistic estimation method for distinctiveness in texts”
  - “Optimal model selection in unsupervised clustering of questions and answers in political discourse”

*Committee:* Sam Bowman, Kyunghyun Cho, Jonathan Nagler, Arthur Spirling (chair), Brandon Stewart

**New York University** | MPhil in Data Science, 2020

- Awarded upon advancement to PhD candidacy

**New York University** | MA in Politics, 2018

- Fully funded Politics PhD student, 2015–2018 (adviser: Arthur Spirling)
- PhD Qualifying Paper: “Estimating the distribution of state power at the UN using topic models”
- Masters’ Paper: “Modelling responsiveness and willingness to negotiate using sentiment: a textual analysis of the FARC-Colombia peace talks”

**Columbia University** | BA in English Literature and Political Science, 2011

- Graduated *summa cum laude* with 3.99 GPA
- Dean’s List every semester. Phi Beta Kappa honor society

## PUBLICATIONS

**Huang, L.**, P. Perry and A. Spirling. “A General Model of Author ‘Style’ with Application to the UK House of Commons, 1935–2018.” *Political Analysis* 28(3):412–434, 2020.

[doi.org/10.1017/pan.2019.49](https://doi.org/10.1017/pan.2019.49)

**Huang, L.**, K. Cho, and S. Bowman. “Can we identify word senses from deep contextualized word embeddings without supervision?” [Link to Medium article.](#)

## PROFESSIONAL EXPERIENCE

**Bloomberg LP** | Research Intern, AI

June–September 2020

- Build and train language models and deep learning models to improve classifier performance on a Terminal product. Technologies used include PyTorch, Transformers and other state-of-the-art architectures, Hadoop, and distributed GPU training.

**RStudio** | Engineering Intern, Shiny team

June–August 2019

- Developed software to enhance the shiny ecosystem maintained by RStudio, the leading IDE for data science in R. Deliverables include the `leaflet.providers` package and a reactive web app.

## **NYU, Center for Data Science** | “NLP/Text-as-Data Speaker Series” Coordinator

June 2017–May 2020

- Manage logistics and communications for weekly guest talks by cutting-edge NLP researchers from top universities and companies such as Google, Facebook, and Microsoft Research ([website](#)).

## **NYU, Politics** | Research Assistant

June–September 2017

- Wrote/refactored simulations in R for a study on incomplete sampling of networks in experiments.

## **NYU, Environmental Studies** | Research Assistant

January–August 2017

- Implemented unsupervised machine learning models. Cleaned and constructed dataset of webscraped texts.

## **W. W. Norton & Company** | Assistant Managing Editor

February 2013–July 2015

- *Promoted from Project Editor in September 2013.*
- Ensured accuracy of all text and visual elements during the publication process from manuscript to finished book, in coordination with authors and production/editorial staff.

## **Foreign Policy Association** | Assistant Editor

June 2011–February 2013

- Guided leading scholars in developing book chapters on international affairs. Wrote and researched supplementary publications. Designed print/digital publications. Hired/supervised interns.

## TEACHING

### **NYU, Center for Data Science** | Teaching Assistant for PhD/MS courses

- Text-as-Data (R), Spring 2020
- Intro to Data Science (Python), Fall 2018
- Text-as-Data (R), Spring 2018

## SOFTWARE

### **R packages on CRAN**

- **stylest**: Estimates a new stylometric measure for textual similarity, especially designed to enable comparisons across large groups of authors/speakers.
- **zipR**: Implements Pythonic `zip()` with convenient options to repeat, broadcast, and fill arrays.
- **leaflet.providers**: Enables users to get and retrieve data from providers of map tiles for the leading interactive map visualization package `leaflet`.

## AWARDS AND FELLOWSHIPS

- Center for Data Science Fellowship, 2018–present
- Society for Political Methodology Travel Grant, 2018, 2019
- Center for Data Science Supplementary Fellowship Grant, 2018
- American Political Science Association Travel Grant, 2018
- NYU Dept. of Politics Summer Research Funding, 2016, 2017, 2018
- MacCracken Fellowship, 2015–2018

## CONFERENCE PRESENTATIONS

- Talk; Natural Language, Dialog and Speech Symposium, 2019
- Poster; Society for Political Methodology Annual Meeting, 2019
- Paper presentation; Conference on Data Science and Law, 2019
- Paper presentation; American Political Science Association Annual Meeting, 2018
- Paper presentation; Society for Political Methodology Annual Meeting, 2018
- Poster; Society for Political Methodology Annual Meeting, 2018
- Invited participant; Text Analysis Developers' Workshop, 2018

## SKILLS

### **Programming**

Python

R

SQL (SQLite, MariaDB)

Spark (PySpark)

### **Methods**

Machine learning (sklearn, pandas)

Deep learning (PyTorch, transformers)

Text mining (spacy, nltk)

Data visualization

Causal inference