

SALLY GAO

github.com/sally-gao | Boston, MA

EXPERIENCE

Thomson Reuters

Data Scientist at Thomson Reuters Innovation Labs | Boston, MA

Aug 2018 – Present

- Develop new data science products and proofs-of-concept for internal and external clients
- Pursue research avenues and develop solutions using machine learning, deep learning and NLP techniques
- Work as part of a cutting-edge team with a focus on innovation and experimentation

The Culture Trip

Hong Kong Editor | Hong Kong

Aug 2016 – Jun 2017

- Covered culture, travel and news in Hong Kong and mainland China
- Wrote and researched ~30 stories per month; regularly conducted interviews with local artists

Time Out Hong Kong

Staff Writer | Hong Kong

May 2016 – Oct 2016

- Reported and wrote feature articles focused on dining, arts and local issues for print and digital publications
- Served as liaison and coordinator between local businesses and magazine's design team

Mental_floss

Freelance Writer & Researcher | New York, NY

Sep 2015 – Apr 2016

- Pitched and authored stories for a print magazine with a circulation of 150,000
- Provided in-depth research and ideas related to hard-to-find stories, untold histories and esoteric facts

EDUCATION

University of Virginia, Data Science Institute

Master of Science in Data Science | GPA: 3.88 | May 2018

Columbia University, Columbia College

Bachelor of Arts in English, *cum laude* | GPA: 3.83 | May 2015

PROJECTS

- **Visualizing crowds in the era of discontent:**

An interactive data visualization showing protests and demonstrations in the U.S. throughout 2017.

link: sallygao.net/crowdviz/crowds.html

- **Exploring recipe ingredients with Latent Dirichlet Allocation:**

Algorithmically discovered ingredient combinations in recipe data in a novel application of topic modeling.

blog: bit.ly/2HNRS1i | **code:** github.com/sally-gao/recipes

- **Analyzing opioid abuse treatment completion with multilevel modeling:**

Built a multilevel model (N = 580,836) to analyze opioid treatment completion at the state and patient level.

Technical paper, poster and presentation presented at the Systems & Information Engineering Design

Symposium (SIEDS '18); winner of the Best Poster Award

SKILLS

Machine learning, Natural language processing (NLP), Statistical modeling, Data visualization, Python, R, SQL, HTML/CSS, Javascript/D3.js