SALLY GAO

github.com/sally-gao · Charlottesville, VA

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 (Dean's List) | May 2015

KNOWLEDGE & SKILLS Machine learning algorithms · Statistical modeling · Natural language processing · Bayesian inference · Python · R · Relational algebra · HTML/CSS · Javascript/D3.js

DATA SCIENCE PROJECTS

Classifying textual styles in natural language Apr – May 2018

- Built a model that matches a writer's individual writing style to the styles of different publications
- Used exclusively topic-independent statistical features, including POS tag n-grams
- Successfully demonstrated early stage proof-of-concept, with promising results from XGBoost and SVM
- ***code**: github.com/sally-gao/styleclassifier_tm*

Visualizing crowds in the era of discontent Apr – May 2018

- Created an interactive visualization of protests, marches and demonstrations in the U.S. throughout 2017
- Cleaned and pre-processed data in Pandas; used HTML and D3 to realize original bubble timeline concept
- ***link:** sallygao.net/crowdviz/crowds.html* · ***code**: github.com/sally-gao/crowdviz*

Exploring recipe ingredients with Latent Dirichlet Allocation Dec 2017 – Jan 2018

- Found common ingredient combinations in recipe data in a novel application of topic modelling
- 5000+ recipes scraped using Python; processing and modelling done in R
- Invented original method of transforming recipes into "bags of words"-style documents
- ***blog:** bit.ly/2HNRS1i* · ***code**: github.com/sally-gao/recipes*

Analyzing opioid abuse treatment completion with multilevel modelling Oct 2017 – Mar 2018

- Built a multilevel model to explain why opioid misuse patients drop out of treatment programs
- Combined two federal datasets (N = 580,836) to analyze treatment completion at the state and patient level
- Found that greater statewide Vivitrol adoption could potentially improve treatment completion rates
- Paper published at the Systems and Information Engineering Design Symposium (SIEDS '18)

EMPLOYMENT HISTORY

The Culture Trip Hong Kong, China
Hong Kong Editor (Freelance) Aug 2016 – Jun 2017

- Pitch and write articles related to culture, travel and news in Hong Kong and China

Time Out Hong Kong Hong Kong, China
Staff Writer Jun 2016 – Oct 2016

- Wrote feature articles, restaurant reviews and in-depth reporting for major lifestyle publication
- Served as liaison between local businesses and design team

Mental_floss New York, NY
Writer/Researcher (Freelance) Sep 2015 – April 2016

- Pitched and authored stories for a print magazine with a circulation of 150,000