**Zehao Xu**

**Phone**: 1(226)5076676 **Email**: z267xu@uwaterloo.ca **Github**: https://github.com/z267xu

**Personal Website**: https://zehaoxu.netlify.app/

## Skills

- Skills: Data Visualization (`ggplot2`, `shiny`, `dash`), Exploratory Data Analysis, Machine Learning, `ETL` (Extract, Transform, and Load process), Optimization, Statistical Computing, Statistical Test, Simulation, Regression, Time Series, Sampling, Experimental Design.
- Specialization: Statistics, Graphical Systems
- Language: `R` (Expert), `Python`, `C++`, `SQL`
- Methods: ANOVA, Generalized Linear Model, Generalized Linear Mixed Model (`glmm`, `lme4`), Generalized Additive Model (`mgcv`), Text Mining (`n-grams`, `tf-idf`), Natural Language Processing (`word2vec`, `latent dirichlet allocation`, `glove`), K Nearest Neighborhood, Classification and Regression Tree, Random Forest, Xgboost, Neural Network, Deep Learning (`keras`, `tensorflow`), Tidyverse (`ggplot2`, `dplyr`), etc

## Education

- Ph.D. in Statistics — University of Waterloo, 11/2021 (expected)
- M.S. in Computational Math — University of Waterloo 8/2017
- B.S. in Statistics — Southwestern University of Finance and Economics, China 8/2016

## Experience

**Univertsity of Waterloo**(*Ph.D.*) Waterloo, Canada 2016 – 2021

- "Exploring data by interactive visualization: an introduction to the `loon` system" (**author, technical support**, submitted to *WIREs Computational Statistics*)
- "`loon.ggplot`: an `R` package to turn `ggplots` to interactive loon plots and vice versa" (in edition, **sole author**). Package loon.ggplot (**sole author**; published); total download approx 8000.
- ggmulti (**sole author**; published): package `ggmulti` extends the `ggplot2` package to provide some high dimensional visualization tools; total download approx 4300.
- loon.tourr (**sole author**; published): it provides grand tour mechanism to project a high dimensional space to a low dimensional space; total download approx 2300.
- loon.shiny (**sole author**; published): package `loon.shiny` renders interactive plots (`loon`) into a web app (`shiny`); total download approx 2100.
- Mini Projects:
  - Bikeshare Analysis: use machine learning methods to predict the number of bikes taken and returned to each station;
  - Twitter Analysis: use NLP to predict whether a posted twitter announce a disaster.

**Plotly**(*Software Engineering*) Montreal, Canada 5/2019 – 8/2019

- Intern `R` team lead: edit and polish `DashR`.
- Develop new package rasterly (**sole author**; published): for large data (even billion) visualization; total download approx 9500.

**MangoTechEducation**(*Data Analysis Supervisor, part-time*) Canada 8/2017 – 4/2019

- Supervise students to accomplish machine learning projects, published projects are
  - Suicide Prediction Analysis with Generalized Addictive Model: study the suicide rate across the world in the past 50 years with neural network.
  - Machine Learning Prediction of New York Airbnb Prices: predict New York Airbnb price with various machine learning methods (e.g. `xgboost`, `randomForest`, etc)