SIJIA HE

Tel: +1- 2167780566 | **E-mail**: sijiahe957@sina.com

EDUCATION

Case Western Reserve University (CWRU) Cleveland, U.S
Bachelor of Science (Major in Statistics) Aug. 2018-Jun. 2022

- **Major GPA**: 3.813/4.0 **Cumulative GPA**: 3.641/4.0
- **Academic Honors**: ***Dean's High Honors*** in Fall 2019, Spring 2020, and Spring 2021; ***Dean's Honor List*** in Fall 2020
- **Coursework**: Calculus; Multivariable Analysis & Data Mining; Data Models & Prediction; Discrete Mathematics; Data Analysis & Linear Models; Explanatory Data Science; Data Visualization & Analysis

RESEARCH EXPERIENCE

Research on the Prediction of Cancer Based on Gene Analysis Cleveland, U.S
Research Assistant, Advisor: Hao (Harry) Feng Sep. 2021-Present

- Cleaned and sorted large volumes of real cancer data into different categories
- Analyzed more than100,000 columns of data to select variable test data
- Conducted variable test and matrix analysis
- Wrote code with R programming language
- Selected the abnormal genes that influence the traits of cells

NeuCA Web Server: A Neural Network-based Cell Annotation Tool with Web-app and GUI Cleveland, U.S
Research Assistant/Second Author, Advisor: Hao (Harry) Feng Jun. 2021- Nov. 2021

- Designed and developed a web app for a neural network-based cell annotation tool
- Prepared and used large volumes of data to test and improve the functions and accuracy of the new web app
- Trained the labeled dataset as SingleCellExperiment class with R before uploading it
- Uploaded the prepared file to the web GUI
- Selected the model size, identified the testing object, and clicked the "Generate Prediction Labels" button to start the prediction process
- Downloaded the results using the formats of RData. csv, or .txt.
- Succeeded in providing researchers with a user-friendly web app with GUI to annotate their scRNA-seq data
- ***Achievement***: the final paper was published in Bioinformatics, Oxford University in November 2021

COURSE PROJECT

Analysis of the Four Risk Factors that Contribute to the Rate of Life Expectancy in Five States in the US Cleveland, U.S
Sep. 2021-Nov. 2021
Programming tools used: R Programming

- Selected and cleaned data categories at first
- Created a multi-category variable and a 4-category variable
- Chose adult obesity, food insecurity, life expectancy, frequent mental distress, and adult smoking as our five variables, and life expectancy as the outcome variable
- Analyzed the relationship between food insecurity and life expectancy in the five states, which are Ohio, North Carolina, Florida, California, and Washington
- Set index of food insecurity as the predictor to predict life expectancy (the predictor variables have 352 data, and the outcome variable have 352 data)
- Visualized the data and built a linear regression model
- Fit new regression model using the Box-Cox transformation, and conducted residual analysis
- Researched the relationship between the level of mental distress frequency and the highest mean level of life expectancy in 352 counties in the five states
- Visualized the data, and built and fitted a linear model,
- Did AVOVA comparisons and predication analyses and concluded that the lowest level of mental distress frequency is associated with the highest mean level of life expectancy in the selected states

OTHER EXPERIENCE

Teaching Experience at Case Western Reserve University Cleveland, U.S
Teaching Assistant for Professor Mathew Williams Feb. 2020-Present

- Assisted more than 180 undergraduate students in learning Matlab
- Worked 9 hours a week to answer questions and grade students' homework and papers during office hours

SKILLS/STRENGTHS

- **Technical Skills**: Matlab; R Programming; SAS
- **Language Skills**: Chinese (Native Speaker); English (Fluent)
- **Strengths**: Cooking; Bodybuilding; Handcrafting Jewelry