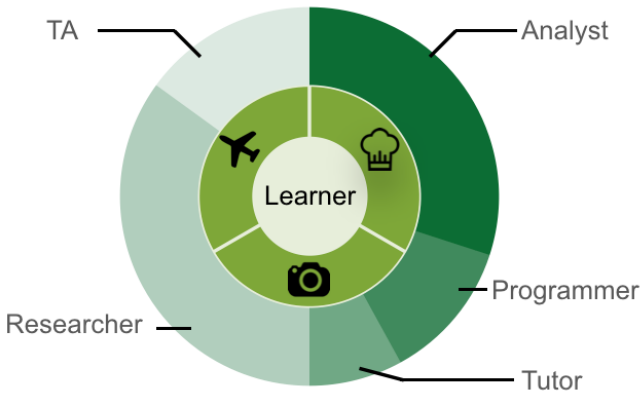
Jianru Shi

sjru123@gmail.com
352-226-6569

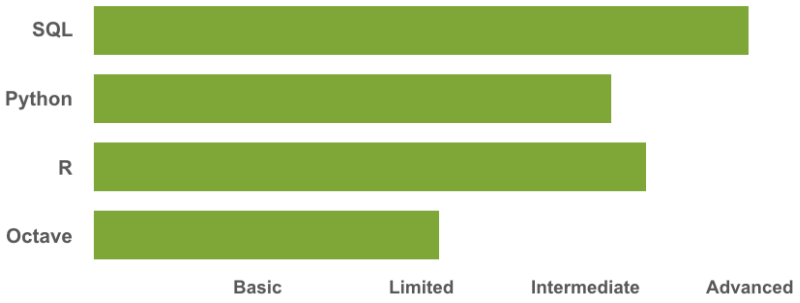
I'm a



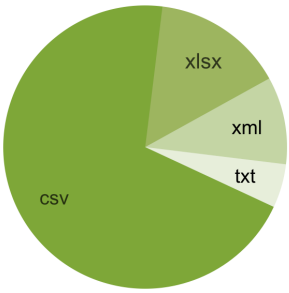
I like

Skills

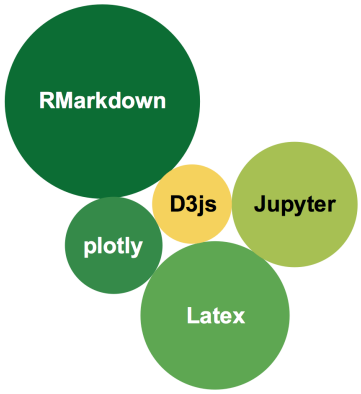
Programming



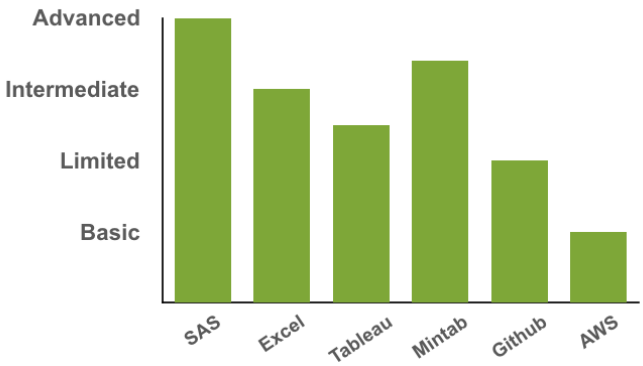
Data Types



Reporting Tools



Analytics Tools



https://www.linkedin.com/in/jianrushi
https://jianru-shi.github.io/

Jianru Shi

Newark, CA, 94560
sjru123@gmail.com
352-226-6569

Summary

Highly motivated graduate student in Statistics with 5 years of hands-on data analysis experience pursuing a position as Data Analyst.
Computer Skills: SAS, Python, SQL, R, Minitab, Excel, Octave
Certifications: SAS BASE programmer for SAS 9.0, SAS Advanced Programmer for SAS 9.0

Education

MS, Statistics, California State University - East Bay (GPA 4.0/4.0) Sep 2015 - Present
PHD, Soil and Water Science, University of Florida (quit with MS) Jun 2015
MS, Natural Resources, University of Nebraska - Lincoln Aug 2013
BS, Environment Science, Nanjing Agricultural University Jun 2011

School Courses: Bayesian Statistics, Probability Theory, Statistical Method in Research, Advanced Statistical Inference, Environmental Statistics, Stochastic Processes, Regression Theory, R for Statistical Learning, Advanced SAS Programing, Mathematical Statistics, Advanced Probability, Data Visualization, Business Communication

MCCO Courses: R programming, The Data Scientist's Toolbox, Getting and Cleaning Data, Exploratory Data Analysis, Learn to Program - the Fundamental, Machine Learning, Data Analyst Nano Degree (Udacity)

Projects

Baseball Players Analysis Jan 2017-Feb 2017

- Worked on a database contains statistics of all baseball players rom 1871 to 2014.
- Performed exploratory data analysis using matplotlib and plotly in python.
- Conducted Spearman's rank correlation analysis and found the top five statistics related to baseball players' salaries.

Titanic Survival Prediction (Kaggle) Jun 2016-Jul 2016

- Implemented data munging and cleaning operations using the pandas package in Python.
- Analyzed and identified relevant features from the dataset by using plotting functions from the 'matplotlib' library.
- Use decision tree, random forest, and logistic regression models to predict the survival status of passengers, the best predict accuracy is 0.81 using random forest classifier.

Accuracy of Food Calories Labels (course project) Nov 2015-Dec 2015

- Conducted one sample t-test on a published dataset to evaluate the accuracy of food calories label. (Minitab)
- Conducted Kruskal Wallis test to determine the factor that affects the food label accuracy. (Minitab)

US-India Joint Clean Energy Research (master thesis) Aug 2013-Jun 2015

- Collected and cleaned laboratory/field data. (Excel)
- Applied multivariate analysis (ANOVA) to quantify the potential environmental hazard of land application of a bioenergy waste material. (SAS)

Experience

Statistics Tutor, Student Center for Academic Achievement, CSUEB 2017-Present
Teaching/Research Assistant, Soil and Water Department, UFL Aug 2013-Jun 2015
Research Assistant, School of Nature Resources, UNL Jan 2012-Aug 2013

https://www.linkedin.com/in/jianrushi https://jianru-shi.github.io/