Sebastian Sciarra, PhD

Senior Data Scientist

seb@sciarra.io
sebastiansciarra.com
sebsciarra
Sebastian Sciarra

Education

PhD | Industrial-Organizational Psychology
University of Guelph
Sep. 2018–May 2023

MSc | Cognitive Psychology
McMaster University
Sep. 2016–June 2018

Honours BSc | Psychology, Neuroscience & Behaviour
McMaster University
Sep. 2012–June 2016

Selected whitepapers

The Theory, Meaning, and Applications of the Singular Value Decomposition
Published
Upcoming

The Game of Supervised Machine Learning: Understanding the Setup, Players, and Rules
Published
10 August 2023

The Expectation-Maximization Algorithm: A Method for Modelling Mixtures of Distributions
Published
28 April 2023

Probability, Likelihood, and Maximum Likelihood Estimation
Published
19 March 2023

Profile

Passionate about coding, machine learning, and statistics. Completed my PhD dissertation at the intersection of these fields to address a practical problem in Industrial-Organizational psychology and received the 2022/2023 Canadian Psychological Association Certificate of Academic Excellence for this work. In my dissertation, I coded and ran Monte Carlo simulations on an AWS instance to evaluate the performance of nonlinear longitudinal models and wrote an R package (nonlinSimsAnalysis) to automate the cleaning, analysis, and visualization of large data sets. Recently at J.D. Power, I have leveraged the Shiny (for Python) framework to automate the creation of Shiny APIs (Shiny example). Whenever I get a chance, I write white papers on machine learning topics at sebastiansciarra.com.

Skills

Coding languages
- Python
- R
- SQL
- LaTeX
- Javascript
- HTML
- CSS

IDEs/platforms
- VSCode
- RStudio
- AWS
- GitHub
- MySQL

Technical skills
- Data visualization (ggplot2, plotnine, matplotlib, seaborn)
- Dynamic data visualization (Shiny, manim)
- Data cleaning (tidyverse, pandas, numpy)
- Machine learning (e.g., regularized regression, multiple imputation, mixture models, etc.)
- Statistics (e.g., latent variable models, factor analysis, multilevel modelling, multilevel modeling, etc.)

Employment experience

Research Data Scientist **Sep. 2023–Current**
J.D. Power
- Conduct analyses for syndicated studies in healthcare and hospitality industries and convey results to non-technical (internal) stakeholders. Analyses include simpler ones like multiple regression to more complicated ones like factor analysis and MICE imputation.
- Continuously improve data quality. Examples include investigating validity of digital fingerprinting tools that identify fraudulent/duplicate respondents, using longitudinal response patterns to identify poor-quality data, and devising trap questions to identify low-effort respondents.
- Leveraged the Shiny (for Python) framework to create of web-based API versions of Excel-based deliverables. Code automated creation of deliverables, reduced incidence of errors, and enhanced user engagement Shiny example.

Teaching Assistant **Sep. 2018–May 2023**
University of Guelph
- Created R scripts for assignments and taught labs for the following courses in measurement and statistics:
  - PSYC 3290 (Conducting Statistical Analyses in Psychology)
  - PSYC 3250 (Psychological Measurement)
  - PSYC 6060 (Research Design and Statistics)
  - PSYC 6380 (Psychological Applications of Multivariate Analysis)
- Taught a variety of topics in methdods and statistics (e.g., regression with continuous and categorical [i.e., ANOVA] variables, *p* values, *p* hacking, hierarchical linear modelling, factor analysis, latent variable modelling, etc.)

Graduate Research Assistant **Sep. 2020–Apr. 2021**
University of Guelph (Part-Time)
- Used R to clean data, compute descriptive statistics, and run regression analyses (with categorical and/or continuous variables) for organizational data on turnover, downsizing, and growth

Data science experience

smltheory **Aug. 2023**
Python package
- Functions within package (9 modules, 30 functions) simulate data sets and demonstrate propositions of supervised machine learning theories (e.g., bias-variance tradeoff, excess risk decomposition)

cobaltResumePro **Aug. 2023**
R Package
- Automates generation of resumes and cover letters within RStudio
- Updated version of my cobaltResume package that generates resumes according to a more streamlined and professional design
- A template and class file were created (~900 lines of LaTeXcode) to specify a styling template

May 2024
Citizenship: Canadian

Education

PhD | Industrial-Organizational Psychology
University of Guelph
Sep. 2018–May 2023

MSc | Cognitive Psychology
McMaster University
Sep. 2016–June 2018

Honours BSc | Psychology, Neuroscience & Behaviour
McMaster University
Sep. 2012–June 2016

Selected whitepapers

The Theory, Meaning, and Applications of the Singular Value Decomposition
Published
Upcoming

The Game of Supervised Machine Learning: Understanding the Setup, Players, and Rules
Published
10 August 2023

The Expectation-Maximization Algorithm: A Method for Modelling Mixtures of Distributions
Published
28 April 2023

Probability, Likelihood, and Maximum Likelihood Estimation
Published
19 March 2023

seb@sciarra.io
sebastiansciarra.com
sebsciarra
Sebastian Sciarra

- R functions were created to easily generate resume entries and merge resume and cover letters into one PDF file

sebastiansciarra.com **Mar. 2023**
Personal website
- Used HTML, JavaScript, and CSS to create a personal website for writing white papers
- White papers focus on statistics, machine learning, and coding by explaining technical details, providing demonstrations, and conducting simulation experiments
- White papers use code from a variety of languages to explain content. As an example, my post titled "Coding and Visualizing the Expectation-Maximization Algorithm" used R, Python, and CSS code

guelphdown **Mar. 2023**
R Package
- Package that automates the generation of theses according to the University of Guelph formatting requirements
- A template and class file were created (~1400 lines of LaTeXcode) to specify formattings for the preamble, body, references, and appendices
- An example of the formatting can be seen in my thesis

nonlinSimsAnalysis **Mar. 2022**
R Package
- Package comprising 105 functions that automate the cleaning, analysis, and visualization of large data sets (e.g., 40 000+ rows) for my doctoral dissertation.
- Common procedures automated by package include combinations of filtering, joining, grouping, aggregating, etc.
- Some examples of figures and tables produced by package: Figure 1, Figure 2, and Table 1.

nonlinSims **Jan. 2022**
R Package
- Package of 30 functions for running simulation experiments of my doctoral dissertation
- The peformance of nonlinear longitudinal models (e.g., structured latent growth curve models) are evaluated under several conditions

Learning SQL **Mar. 2021**
Project
- Completed through 16 of 18 chapters from Alan Beaulieu's Learning SQL
- Topics include filtering, querying multiple tables, sets, grouping and aggregates, subqueries, joins, transactions etc.

May 2024
Citizenship: Canadian