

Vanessa Scott

☎ (858) 255-9128

✉ vleescott333@gmail.com

📍 San Diego, CA

EXECUTIVE SUMMARY

Data scientist with **7+ years of experience** in research, data management, and applied analytics across clinical, behavioral, and digital health settings. Experience in combining **storytelling, predictive modelling, and operational impact** for optimizing business impact through achievements like:

- **Directing data operations** for 3 longitudinal studies (\$6M+, 500+ participants), advancing clinical engagement (>80% participant adherence) and high-risk population monitoring through mobile health data.
- **Building scalable data pipelines** for National Data Archive submissions and harmonized 60+ behavioral and clinical measures (**10+ GB**) using **Python and SQL**, reducing submission time by **75%**.
- **Built real-time dashboards** and curated **data visualizations** to track adherence, dropout, and engagement, identifying retention risks and ensuring alignment with grant and compliance milestones.

EXPERIENCE

UCSD Health Psychiatry Cognitive Dynamics Lab, Dr. Depp & Dr. Moore -- UC San Diego Health, San Diego's sole academic health system, integrates data-driven research and clinical operations, managing \$686M+ in annual health sciences research to enhance over 240K patients care and health systems performance.

Staff Research Supervisor

June 2023 - Present

- **Directed data operations** for 3 longitudinal studies (\$6M+, 500+ participants), advancing clinical engagement (>80% participant adherence) and high-risk behavioral health population monitoring via mobile health data.
- **Engineered multimodal data pipelines** (GPS, audio, facial encoding) and co-developed a domain-specific **NLP RAG LLM** for novel psychiatric feature extraction.
- **Built real-time dashboards** and curated **data visualizations** to track adherence, dropout, and engagement, identifying retention risks and ensuring alignment with grant and compliance milestones.
- **Developed scalable data infrastructure** using **REDCap API, AWS S3**, and automated logic tools, improving data integrity and **operational efficiency for over 20 researchers, 500+ users on 100+ GB of data**.
- **Led and trained** an 8+ member team to standardize behavioral data workflows and ensure privacy-compliant, reproducible analytics practices.

Staff Research Analytics Scientist II

July 2022 - June 2023

- **Built scalable data pipelines** for National Data Archive submissions and harmonized 60+ behavioral and clinical measures (**10+ GB**) using **Python and SQL**, reducing submission time by **75%**.
- **Performed exploratory/inferential analyses** (regression, correlation) using Python and SPSS to derive insights aimed at **identifying high-risk individuals** and **user behavioral patterns**.
- **Digitized legacy paper-based data collection** with CRUD-enabled forms, improving data accuracy and reducing human **data entry errors up to 100%**.
- **Automated digital recruitment and screening systems**, increasing participant leads by **300%** and improving retention and engagement by **50%**.

Staff Research Analytics Scientist I

Aug. 2021 - July 2022

- **Queried and engineered large-scale datasets (30+ GB)** for robust mobile text-survey data from **AWS** using **SQL** to extract features for downstream analytics and modeling.
- **Streamlined data processing** (cleaning, merging, transformation) with **pandas and NumPy**, reducing manual workload, mitigating previous human-error protocols and accelerating analyses.
- **Developed IRB-compliant research protocols and recruitment workflows**, ensuring ethical, high-quality **HIPAA-compliant** clinical data collection.

UCR Perception, Action, and Development (PAD) Lab, Dr. Franchak -- The UCR PAD Lab has diverse-funded research projects investigating motor and cognitive development in infants using wearable sensors, NLP devices, machine learning models and high frequency mobile surveying.

Research Project Manager

Aug. 2019 - Aug. 2021

- **Co-authored a published ML study** using a **random forest model** to predict infant movement with up to **95% accuracy**, pioneering a novel contactless behavioral tracking study paradigm.
- **Designed and deployed ecological momentary assessment (EMA) studies** for remote behavioral monitoring during COVID-19, adapting legacy protocols for digital data capture.
- **Directed NIH-funded Play Project operations**, developing SOPs, training staff, and overseeing custom data collection infrastructure.
- **Programmed in MATLAB and R** for behavioral coding, signal processing and statistical modeling.

UCR Brain Game Center Research Lab, Dr. Seitz -- UCR Brain Game Center where I worked on NIMH-funded project conducting large-scale video game-based cognitive training studies with 30,000+ participants.

Research Project Manager

June 2017 - June 2018

- **Led research of gamified cognitive training paradigms and trained and supervised 10+ research assistants** to ensure consistent, high-fidelity experimental data collection.

SKILLS

Programming Languages & Tools (3+ years):

Python (pandas, NumPy, SciPy, scikit-learn, spaCy, NLTK, Spark, etc.) SQL, SPSS, REDCap

Data Science & Machine Learning (3+ years):

Predictive/Bayesian Modeling, Neural Net., Clustering & Regression Methods, XGBoost/Gradient Desc. Techniques

Data Engineering, Visualization & Management (7+ years):

ETL, RDBMS, AWS, Git, V.S. Code, HTML/CSS/Java Script, Looker, HIPAA-Compliant Systems

Research & Project Management (7+ years):

Digital Health Research, National Data Reporting (e.g., NIMH, NDA), Scientific Writing/Presentation

EDUCATION

UC San Diego, Halicioğlu Data Science Institute – M.S. Data Science (Sept 2024 – Apr 2026)

**This is a career-professional program, I am available for full-time roles now.*

UC Riverside – B.S. Psychology, Research Design (Sept 2014 – Jun 2018)

Mt. San Jacinto College – Leadership Management (Jan 2019 – May 2019)

San Diego Mesa College – Python Coding & Python for Data Science (Sept 2022 – Jun 2023)