

# Tinashe Michael Tapera

RESEARCH SOFTWARE ENGINEER AT HARVARD UNIVERSITY

Harvard University, Boston, MA

+267 441 7206 | [tinashemtamera@gmail.com](mailto:tinashemtamera@gmail.com) | [tinashemtamera.com](https://tinashemtamera.com) | [TinasheMTapera](https://www.linkedin.com/company/tinashemtamera) | [TinasheMTapera](https://www.linkedin.com/company/tinashemtamera) | [TaperaTinashe](https://twitter.com/TaperaTinashe)

*I'm a Research Software Engineer at Harvard University in the Golden Planetary Health Lab and National Studies on Air Pollution & Health (NSAPH). I write software that makes science happen by developing high-throughput ETL data pipelines, distributing accessible research software packages, and providing documentation and support for best practices in data science at Harvard.*

## Education

### Northeastern University

MSC HEALTH INFORMATICS

- Capstone: TBD

Boston, MA

Sep 2022 — May 2026

### Drexel University

ACCELERATED MSC. PSYCHOLOGY (DATA ANALYSIS TRACK)

- Thesis: Advanced Data Mining Methods for Psychological & Behavioral Research
- GPA: 3.70 (cum laude)

Philadelphia, PA

Sep 2017 — Jun 2018

### Drexel University

BSC. PSYCHOLOGY

- GPA: 3.52
- A.J. Drexel Scholarship
- Dean's List

Philadelphia, PA

Sep 2013 — Jun 2017

## Professional Experience

### Golden Planetary Health Lab, National Studies on Air Pollution & Health, Harvard University

RESEARCH SOFTWARE ENGINEER

- Support interdisciplinary research at the intersection of climate science, public health, and data science.
- Develop scalable, reproducible software solutions for high throughput data ETL and harmonization pipelines from various diverse sources
- Deploy robust and well-documented software packages to seamlessly interface with small- to large-scale harmonized data sources

Boston, MA

Feb 2025 — Present

### ConcertAI

DATA SCIENCE INTERN

- Delivered Real World Data (RWD) oncology data products leveraging curated data, unstructured EMR, and claims records
- Deployed R Shiny Dashboard for pharmacovigilance of Adverse Drug Events related to Multiple Myeloma treatment
- Analyzed Adverse Events using the Tree-Based Scan Statistic data mining method

Boston, MA

May 2024 — Oct 2024

### Penn Lifespan Informatics & Neuroimaging Center

SENIOR NEUROIMAGING DATA ANALYST

- Developed data pipelines for ETL and analysis of large-scale neuroimaging data sets between data warehouses in Python, R, and Bash
- Preprocessed and analysed neuroimaging data using cutting-edge software (fMRIPrep, XCPEngine, QSIprep, ASLPrep)
- Maintained and supported multiple data curation software packages in Python and R

Philadelphia, PA

Oct 2018 — Aug 2022

### Salesforce

DATA SCIENCE INTERN

- Focused on discovery of organizational insight using internal human resources data sets
- Developed a semi-supervised learning algorithm to track employee performance by matching topic models of continuous feedback and goal-setting data
- Investigated comorbidity of employees' insurance claims data to dynamically classify claim types and employee phenotypes

San Francisco, CA

May 2017 — Oct 2017

## Research Experience

### UbiWell Lab (Varun Mishra, PhD)

Northeastern University

“STRESSFREE: ASSESSING THE SCALABILITY & FEASIBILITY OF DIGITALLY PHENOTYPING STRESS”

2022

- Aim: Developing tools and software to identify moments of heightened stress in Northeastern undergraduate students, with the long term goal of delivering just-in-time interventions to relieve stress with mobile-CBT approaches

### Quantitative Psychology & Statistics Lab (Fengqing Zhang, PhD)

Drexel University

“MOMENTARY CHANGES IN HEART RATE VARIABILITY CAN DETECT RISK FOR EMOTIONAL EATING EPISODES.”

2015 — 2019

- Aim: predicting emotional eating episodes in disordered eating patients using a combination of heart rate variability data and self-report

### Quantitative Psychology & Statistics Lab (Fengqing Zhang, PhD)

Drexel University

“APPLICATION OF ADVANCED DATA MINING MODELS TO IDENTIFY DIETARY PATTERNS ASSOCIATED WITH RISK OF

CARDIOVASCULAR DISEASE.”

2015 — 2019

- Aim: compare the performance of unsupervised feature selection (PCA/FA) against regularization (L1/L2) in predicting cardiovascular disease biomarkers from high-dimensional food and behaviour survey responses

### Quantitative Psychology & Statistics Lab (Fengqing Zhang, PhD)

Drexel University

“IMPROVED MODELLING OF SMARTPHONE-BASED ECOLOGICAL MOMENTARY ASSESSMENT DATA FOR DIETARY LAPSE

PREDICTION.”

2015 — 2019

- Aim: predicting dietary adherence lapses in participants using self-reported EMA

### Quantitative Psychology & Statistics Lab (Fengqing Zhang, PhD)

Drexel University

“IDENTIFYING AUTISM DIAGNOSTIC INTERVIEW: REVISED ALGORITHM ITEMS THAT SIGNIFICANTLY DISTINGUISH AUTISM

SPECTRUM DISORDER AND DOWN SYNDROME.”

2015 — 2019

- Aim: Identify phenotypic differences between children with autism spectrum disorder, down syndrome, and comorbid diagnoses using the Autism Diagnostic Interview-Revised (ADI-R)

### Quantitative Psychology & Statistics Lab (Fengqing Zhang, PhD)

Drexel University

“MODELING ZERO-INFLATED MVPA BOUTS USING A HIERARCHICAL LINEAR MODELING FRAMEWORK”

2015 — 2019

- Aim: predict participants' moderate-to-vigorous physical activity (MVPA) bouts at timepoint 3 from previous timepoints using a zero-inflated Tweedie Poisson regression model in a growth curve modeling context

### Statistical and Applied Mathematical Sciences Institute (SAMSI)

NC State University

“PREDICTING MULTIPLE SCLEROSIS (MS)”

2016

- Aim: classify participant diagnosis (MS patient vs. control) using lesion count along the corpus callosum in a diffusion dataset

### Laboratory for Innovations in Health-Related Behavior Change (Evan Forman, PhD)

Drexel University

“A COMPANION SMARTPHONE APP TO ENHANCE DIETARY ADHERENCE THROUGH PREDICTIVE MACHINE LEARNING”

2015

- Aim: Data collection, cleaning, and summarization with Excel and SPSS

## Publications

*First-author*

### Flywheeltools: data curation and manipulation on the flywheel platform

Frontiers in neuroinformatics

TM TAPERA, M CIESLAK, M BERTOLERO, A ADEBIMPE, GK AGUIRRE, ER BUTLER, ...

2021

### DOES ECOLOGICAL MOMENTARY ASSESSMENT DATA REFLECT BASELINE SELF-REPORT IN WEIGHT LOSS TREATMENT?

ANNALS OF BEHAVIORAL MEDICINE

TM TAPERA, S GOLDSTEIN, BC EVANS, E FORMAN

2016

*Middle-author*

### Reproducible Brain Charts: An open data resource for mapping brain development and its associations with mental health

Neuron

G SHAFIEI, NB ESPER, MS HOFFMANN, L AI, AA CHEN, J CLUCE, S COVITZ, ...

2025

### Functional connectivity development along the sensorimotor-association axis enhances the cortical hierarchy

Nature communications

AC LUO, VJ SYDNOR, A PINES, B LARSEN, AF ALEXANDER-BLOCH, M CIESLAK, ...

2024

<b>Diffusion MRI head motion correction methods are highly accurate but impacted by denoising and sampling scheme</b>	<i>Human Brain Mapping</i>
M CIESLAK, PA COOK, G SHAFIEI, TM TAPERA, H RADHAKRISHNAN, M ELLIOTT, ...	2024
<b>Development of top-down cortical propagations in youth</b>	<i>Neuron</i>
A PINES, AS KELLER, B LARSEN, M BERTOLERO, A ASHOURVAN, DS BASSETT, ...	2023
<b>Development of white matter fiber covariance networks supports executive function in youth</b>	<i>Cell reports</i>
J BAGAUTDINOVA, J BOURQUE, VJ SYDNOR, M CIESLAK, AF ALEXANDER-BLOCH, ...	2023
<b>ModelArray: An R package for statistical analysis of fixel-wise data</b>	<i>NeuroImage</i>
C ZHAO, TM TAPERA, J BAGAUTDINOVA, J BOURQUE, S COVITZ, RE GUR, ...	2023
<b>Developmental coupling of cerebral blood flow and fMRI fluctuations in youth</b>	<i>Cell Reports</i>
EB BALLER, AM VALCARCEL, A ADEBIMPE, A ALEXANDER-BLOCH, Z CUI, RC GUR, ...	2022
<b>ASLPrep: a platform for processing of arterial spin labeled MRI and quantification of regional brain perfusion</b>	<i>Nature methods</i>
A ADEBIMPE, M BERTOLERO, S DOLUI, M CIESLAK, K MURTHA, EB BALLER, ...	2022
<b>Curation of BIDS (CuBIDS): A workflow and software package for streamlining reproducible curation of large BIDS datasets</b>	<i>NeuroImage</i>
S COVITZ, TM TAPERA, A ADEBIMPE, AF ALEXANDER-BLOCH, MA BERTOLERO, ...	2022
<b>Mobile footprinting: linking individual distinctiveness in mobility patterns to mood, sleep, and brain functional connectivity</b>	<i>Neuropsychopharmacology</i>
CH XIA, I BARNETT, TM TAPERA, A ADEBIMPE, JT BAKER, DS BASSETT, ...	2022
<b>Spatially-enhanced clusterwise inference for testing and localizing intermodal correspondence</b>	<i>NeuroImage</i>
SM WEINSTEIN, SN VANDEKAR, EB BALLER, D TU, A ADEBIMPE, TM TAPERA, ...	2022
<b>Refinement of Functional Connectivity in Development Aligns with the Sensorimotor to Association Axis</b>	<i>OSF</i>
A LUO, V SYDNOR, AS KELLER, A ALEXANDER-BLOCH, M CIESLAK, S COVITZ, ...	2022
<b>QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data</b>	<i>Nature methods</i>
M CIESLAK, PA COOK, X HE, FC YEH, T DHOLLANDER, A ADEBIMPE, ...	2021
<b>A simple permutation-based test of intermodal correspondence</b>	<i>Human brain mapping</i>
SM WEINSTEIN, SN VANDEKAR, A ADEBIMPE, TM TAPERA, ...	2021
<b>Developmental coupling of cerebral blood flow and fMRI fluctuations in youth</b>	<i>BioRxiv</i>
EB BALLER, AM VALCARCEL, A ADEBIMPE, A ALEXANDER-BLOCH, Z CUI, RC GUR, ...	2021
<b>ASLPrep: a generalizable platform for processing of arterial spin labeled MRI and quantification of regional brain perfusion</b>	<i>BioRxiv</i>
A ADEBIMPE, M BERTOLERO, S DOLUI, M CIESLAK, K MURTHA, EB BALLER, ...	2021
<b>Mobile footprinting: linking individual distinctiveness in mobility patterns to mood, sleep, and brain functional connectivity</b>	<i>BioRxiv</i>
CH XIA, I BARNETT, TM TAPERA, Z CUI, TM MOORE, A ADEBIMPE, ...	2021
<b>Mapping Physiology-Function Coupling in Youth</b>	<i>Biological Psychiatry</i>
E BALLER, A ADEBIMPE, A VALCAREL, A ALEXANDER-BLOCH, Z CUI, J DETRE, ...	2021
<b>Momentary changes in heart rate variability can detect risk for emotional eating episodes</b>	<i>Appetite</i>
AS JUARASCIO, RJ CROCHIERE, TM TAPERA, M PALERMO, F ZHANG	2020
<b>QSIPrep: An integrative platform for preprocessing and reconstructing diffusion MRI</b>	<i>Biorxiv</i>
M CIESLAK, PA COOK, X HE, FC YEH, T DHOLLANDER, A ADEBIMPE, ...	2020
<b>Autism spectrum disorder (ASD) symptom profiles of children with comorbid Down syndrome (DS) and ASD: A comparison with children with DS-only and ASD-only</b>	<i>Research in Developmental Disabilities</i>
M GODFREY, S HEPBURN, DJ FIDLER, T TAPERA, F ZHANG, CR ROSENBERG, ...	2019

Application of a new dietary pattern analysis method in nutritional epidemiology

F ZHANG, TM TAPERA, J GOU

A PRELIMINARY INVESTIGATION OF A PERSONALIZED RISK ALERT SYSTEM FOR WEIGHT CONTROL LAPSES

E FORMAN, S GOLDSTEIN, B EVANS, S MANASSE, A JUARASCIO, M BUTRYN, ...

IS PROMPTING PROBLEMATIC?: CONSIDERATIONS FOR LONG-TERM ECOLOGICAL MOMENTARY ASSESSMENT

SP GOLDSTEIN, BC EVANS, TM TAPERA, E FORMAN, S MANASSE, A JUARASCIO, ...

BMC medical research methodology

2018

ANNALS OF BEHAVIORAL MEDICINE

2016

ANNALS OF BEHAVIORAL MEDICINE

2016

## Software & Project Contributions

FlywheelTools

A SUITE OF SOFTWARE TOOLS FOR CURATING YOUR NEUROIMAGING DATA INTO BIDS ON FLYWHEEL

R, Python, MongoDB, RMarkdown Reports, Docker

10.5281/zenodo.4752798

2021

PC Dashboard

AN INTERACTIVE ANALYTICS DASHBOARD FOR THE DREXEL UNIVERSITY PEER COUNSELING HELPLINE

R, Shiny, Qualtrics API

10.5281/zenodo.5721127

2017

## Teaching Experience

Teaching Assistant	
2018	MSc. Psychology — Statistics I & II
	Drexel University

## Service

2022 to present	Member, Mentor	R 4 Data Science Community (R4DS.io)
2025 to present	Newsletter Editor	US Research Software Engineering Association
2013 to 2021	Member, Alumni Mentor	Drexel University Gospel Choir
2013 to 2018	Peer Counselor, VP of Scheduling & Communications	Drexel University Peer Counseling Helpline

## Skills

### Analytical

DATA SCIENCE, STATISTICAL MODELLING, REPRODUCIBLE RESEARCH, PARAMETERIZED & INTERACTIVE REPORTS, PLOTTING & VISUALISATION, OBJECT-ORIENTED PROGRAMMING, NOTEBOOK-DRIVEN DEVELOPMENT

### Programming languages

R, PYTHON, BASH

### Databases

MySQL, MONGODB

### Packages

TIDYVERSE, RMARKDOWN/QUARTO, GGLOT2, SHINY, TIDYMODELS, GITHUB PAGES, NILEARN

### Tools

GIT, DOCKER, SINGULARITY, RSTUDIO, POSITRON, VSCODE, JUPYTER, CIRCLECI