

# Philippe Boileau

PHD STUDENT · BIOSTATISTICS

University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA

✉ philippe\_boileau@berkeley.edu | 🏠 pboileau.ca | 📷 PhilBoileau

## Education

---

### University of California, Berkeley

Berkeley, CA

PHD BIOSTATISTICS

2020 - present

- Advisor: Sandrine Dudoit
- Thesis: High Dimensional Covariance Matrix Estimation for Unsupervised Learning Applications

### University of California, Berkeley

Berkeley, CA

MA BIOSTATISTICS

2018 - 2020

- Committee: Sandrine Dudoit, chair; Elizabeth Purdom; Haiyan Huang
- Thesis: Exploring High-Dimensional Biological Data with Sparse Contrastive Principal Component Analysis

### Concordia University

Montréal, QC

BS HONOURS STATISTICS

2016 - 2018

- Advisors: Lisa Kakinami; Lea Popovic
- Honours thesis: Modelling the Obesity Epidemic with Networks

## Professional Experience

---

- 2020-pres. **Graduate Student Researcher**, University of California, Berkeley Superfund Research Program  
Analyse data collected by the organization's environmental health scientists and epidemiologists, applying and developing statistical methods to glean insights on the effects of chemical exposures on human health.
- 2020-pres. **Statistician**, Ha Lab, University of California, San Francisco  
Design experiments, develop bioinformatic pipelines, and apply biostatistical methods for the analysis of bulk transcriptomic data of salivary gland tissue taken from adenoid cystic carcinoma patients.
- 2020 **Statistician**, He Lab, University of California, Berkeley  
Created and applied bioinformatic and biostatistical pipelines for the analysis of single-cell transcriptomic data of lung epithelial cells cultivated from mice.
- 2019 **Graduate Summer Intern**, Sutter Health Research, Development and Dissemination  
Developed a statistical learning pipeline to evaluate a patient's risk of becoming septic during their hospital visit. The model's preliminary results were demonstrably superior to existing methods.
- 2015-2018 **Research Assistant**, Canadian Longitudinal Study on Aging, Statistical Analysis Center  
Worked with biostatisticians, data curators and data access managers on various projects, including the development of an R package to facilitate analysis of proprietary data, the creation of software to verify accuracy of survey responses, and the creation of a public-use dataset.
- 2017 **Undergraduate Summer Researcher**, Concordia University  
Worked under Dr. Lea Popovic and Dr. Lisa Kakinami with the Quebec Adipose and Lifestyle Investigation in Youth cohort data to develop an R package that uses heatmaps to explore high-dimensional ego network data. This package simplifies the exploratory data analysis process for high-dimensional, multi-network data sets through the use of hierarchical clustering, consensus clustering, and interactive heatmaps.

## Publications

---

### PUBLISHED

#### STATISTICAL METHODOLOGY

**Boileau P**, Hejazi N, Dudoit S. “Exploring High-Dimensional Biological Data with Sparse Contrastive Principal Component Analysis,” *Bioinformatics*, vol. 36, no. 11, p. 3422-3430, June 2020. doi: 10.1093/bioinformatics/btaa176.

#### REFEREED STATISTICAL SOFTWARE

**Boileau P**, Hejazi N, Dudoit S. “scPCA: A toolbox for sparse contrastive principal component analysis in R,” *The Journal of Open Source Software*, vol. 5, no. 46, p. 2079, February 2020. doi:10.21105/joss.02079. (Available online: <https://doi.org/10.21105/joss.02079>)

### IN PREPARATION

**Boileau P**, Hejazi N, van der Laan, Dudoit S. “Data-Adaptive Covariance Matrix Estimation in High Dimensions,” 2020.

**Boileau P**, Hejazi N, van der Laan, Dudoit S. “Contrastive Covariance Matrix Estimation: A Frequentist Approach,” 2020.

Cevalos J, **Boileau P**, Ha P. “Evaluation of Fusion Transcripts and the Anti-Tumor Immunity in Salivary Gland ACC and ACC Patient Derived Xenograft Models,” 2021.

### BOOK CHAPTERS

Lun ATL, **Boileau P**. “Case Study: Messmer Human ESC” in “Orchestrating Single-Cell Analysis with Bioconductor,” 2020+, in preparation. (Available online: <http://bioconductor.org/books/devel/OSCA/messmer-hesc.html>)

### TECHNICAL REPORTS

**Boileau P**, Popovic L, Barnett TA, Henderson M, Kakinami L. “Exploration of Multi-Network Data with Heatmaps,” ISM Internship Proceedings, Summer 2017. (Available online: <http://ism.uqam.ca/~ism/pdf/Boileau-scientific%20report.pdf>)

### THESES

**Boileau P**. “Exploring High-Dimensional Biological Data with Sparse Contrastive Principal Component Analysis,” Master’s thesis, University of California, Berkeley, 2020.

**Boileau P**. “Modelling the Obesity Epidemic with Networks,” Bachelors of Honours Science thesis, Concordia University, 2018.

## Awards, Fellowships, & Grants

---

|      |   |           |
|------|---|-----------|
| 2020 | <b>Biostatistics Block Grant and Non-Resident Student Tuition Award</b> , University of California, Berkeley, Division of Biostatistics | \$ 15,102 |
|      | <b>Bourse de maîtrise en recherche</b> , Fonds de recherche dur Québec, Nature et technologies  | \$ 12,667 |
| 2019 | <b>Biostatistics Block Grant and Non-Resident Student Tuition Award</b> , University of California, Berkeley, Division of Biostatistics | \$ 15,102 |
|      | <b>Bourse de maîtrise en recherche</b> , Fonds de recherche dur Québec, Nature et technologies  | \$ 12,667 |
| 2018 | <b>Biostatistics Block Grant and Non-Resident Student Tuition Award</b> , University of California, Berkeley, Division of Biostatistics | \$ 15,102 |
|      | <b>Best Poster Presentation</b> , Cardiometabolic, Diabetes and Obesity Conference  | \$ 300    |
|      | <b>Best Undergraduate Poster Presentation</b> , Canadian Statistics Student Conference  | \$ 150    |
| 2017 | <b>Undergraduate Summer Scholarship</b> , Institut des Sciences Mathématiques   | \$ 3,750  |
|      | <b>Alumni Association Scholarship</b> , Concordia University  | \$ 2,000  |
|      | <b>Faculty of Arts and Science Scholar</b> , Concordia University   |           |

## Presentations

---

### CONTRIBUTED PRESENTATIONS

**Boileau P**, Hejazi N, Dudoit S. “Sparse Contrastive Principal Component Analysis,” BioC2020: Where Software and Biology Connect, Online, July 2020.

### CONTRIBUTED POSTERS

**Boileau P**, Hejazi N, Dudoit S. “Sparse Contrastive Principal Component Analysis,” March 2020, Berkeley Statistics Annual Research Symposium, Department of Statistics, UC Berkeley. **(Cancelled due to COVID-19)**

Scheel D, Perkova A, Danieles P, **Boileau P**, Kakinami L. “Does survey design information matter? Assessing the impact on population estimates of hypertension in Canada,” 2018 Annual Meeting, Statistical Society of Canada, Montreal, Quebec, June 4, 2018.

**Boileau P**, Popovic L, Barnett TA, Henderson M, Kakinami L. “Ego Network Exploration with Heatmaps: A Case Study on Pediatric Obesity,” Canadian Statistics Student Conference, Montreal, Quebec, June 2, 2018.

**Boileau P**, Popovic L, Barnett TA, Henderson M, Kakinami L. “Ego Network Exploration with Heatmaps: A Case Study on Pediatric Obesity,” PERFORM Conference, Montreal, Quebec, May 17, 2018.

**Boileau P**, Popovic L, Barnett TA, Henderson M, Kakinami L. “L’analyse de reseaux sociaux avec cartes thermiques, une etude de cas avec l’obesite pediatrique,” Quebec Society for lipid, nutrition and metabolism scientific meeting, Magog-Orford, Quebec, February 7-9, 2018.

## Teaching Experience

---

|          |   |                    |
|----------|---|--------------------|
| Su 2020  | <b>Data 8: The Foundations of Data Science</b> , Lecturer                           | <i>UC Berkeley</i> |
| Sp 2020  | <b>PBHLTH 241: Statistical Analysis of Categorical Data</b> , Teaching Assistant    | <i>UC Berkeley</i> |
| Fa 2019  | <b>Data 100: Principles and Techniques of Data Science</b> , Teaching Assistant     | <i>UC Berkeley</i> |
| Sp 2019  | <b>Data 100: Principles and Techniques of Data Science</b> , Teaching Assistant     | <i>UC Berkeley</i> |
| Fa. 2018 | <b>PBHLTH 142: Introduction to Statistics in Public Health</b> , Teaching Assistant | <i>UC Berkeley</i> |

## Mentoring

---

|            |  |                    |
|------------|--|--------------------|
| Su/Fa 2020 | <b>Brian Collica</b> , MA Statistics Student, <i>Statistics Research Apprenticeship Program</i><br>“cvCovEst: An R Package for Covariance Matrix Estimator Selection in High Dimensions” | <i>UC Berkeley</i> |
| Fa 2020    | <b>Jamarcus Liu</b> , Undergraduate, <i>Statistics Research Apprenticeship Program</i><br>“cvCovEst: An R Package for Covariance Matrix Estimator Selection in High Dimensions”          | <i>UC Berkeley</i> |
| Sp 2020    | <b>Star Li</b> , Undergraduate, <i>Statistics Research Apprenticeship Program</i><br>“Recovering Gene Expression Datasets’ True Dimensionality: Does Initialization Matter?”             | <i>UC Berkeley</i> |
| Fa 2020    | <b>David Liu</b> , Undergraduate, <i>Statistics Research Apprenticeship Program</i><br>“Recovering Gene Expression Datasets’ True Dimensionality: Does Initialization Matter?”           | <i>UC Berkeley</i> |