# Overview

Max Turgeon

STAT 7200–Multivariate Statistics

## Course details

- Time: MWF 9:30am–10:20am
- Office: 344 Helen Glass Centre
- Office Hours:
  - Tuesday 9:30am–11am
  - Wednesday 10:30am-12pm
  - Or by appointment
- Course Website: `https://maxturgeon.ca/w20-stat7200`

## Textbook

- There is no required textbook for this class. Some references:
    - Anderson, *An Introduction to Multivariate Statistical Analysis.* Wiley (2003).
    - Muirhead, *Aspects of Multivariate Statistical Theory.* Wiley (2005).
    - Johnson & Wichern, *Applied Multivariate Statistical Analysis.* Prentice Hall (2007).

- Three assignments worth a total of 40% of the final grade
- One midterm (tentatively scheduled February 28 outside of class hours) worth 30% of the final grade
- There is **no** final exam
- There is a class project worth 30% of the final grade

# IN THE CASE OF A FIRE ALARM:

- REMAIN CALM
- IF IT IS SAFE, EVACUATE THE CLASSROOM OR LAB
- GO TO THE CLOSEST FIRE EXIT
- DO NOT USE THE ELEVATORS

IF YOU NEED ASSISTANCE TO EVACUATE THE BUILDING, INFORM YOUR PROFESSOR OR INSTRUCTOR NOW!!!

- IF DURING A BUILDING EVACUATION YOU NEED TO REPORT AN INCIDENT OR A PERSON LEFT BEHIND:
  - CONTACT ONE OF THE BUILDING FIRE WARDENS OR
  - CALL SECURITY SERVICES 204-474-9341

- DO NOT REENTER THE BUILDING UNTIL

THE **"ALL CLEAR"** IS DECLARED BY A FIRE WARDEN, SECURITY SERVICES OR THE FIRE DEPARTMENT

## Course Objectives

- Broad overview of techniques used in multivariate analysis, with emphasis on Multivariate Linear Regression and Principal Component Analysis.
    1. Make decisions on how and when to use the techniques discussed in class;
    2. Apply and assess multivariate methods on real data;
    3. Make sound statistical conclusions based on a multivariate analysis.
- Apply these methods using the R statistical software.

## Tentative topics

- Review of linear algebra and multivariate distributions
- Multivariate normal distribution
- Wishart distribution
- Inference about mean vectors
- Inference about covariance matrices
- **Multivariate linear regression**
- **Principal Component Analysis**
- Factor Analysis
- Canonical Correlation Analysis
- If time permits, advanced topics:
    - E.g. sparse models, graphical models, random matrix theory

## Multivariate Data

- Multivariate data is **everywhere**
  - Multiple measurements collected on a given experimental unit
- Multivariate analysis is concerned with the relationship between those variables
- **Note**: Regression with a single outcome variable is *not* considered multivariate analysis.

# Multivariate Methods

- One- or two-sample inference about multivariate data (think t-test)
- *MANOVA*: Generalization to several populations
- Multivariate Linear Regression: Linear model for multivariate response in terms of a set of covariates

- Principal Component Analysis: Reduce dimension of data by finding directions in data with maximal variance
- *Factor Analysis*: Understand variance in multivariate sample in terms of latent (i.e. unobserved) factors
- *Canonical Correlation Analysis*: Study correlations between two multivariate datasets

## Multivariate Methods (not covered in STAT 7200)

- Methods for longitudinal data (e.g. mixed-effect models or GEEs)
- *Clustering*: Grouping "similar" observations based on their (multivariate) measurement (STAT 4600: Statistical Learning)
- *Classification and Discrimination*: Grouping observations and allocating new units to previously defined classes (STAT 4600: Statistical Learning)
    - The difference between the last two is whether or not we measured a class label for the observations.

## Assessments

- This is a theory course, so we will be proving theorems
- However, the assignments and midterm will also discuss applications
- We will use R, and the datasets will be provided to you.
    - Code for in-class examples will also be provided
- For assignments and course project, students are *strongly* encouraged to use Rmarkdown or knitr.
    - Literate programming
    - Reproducibility