An Introduction to Functional Data Analysis

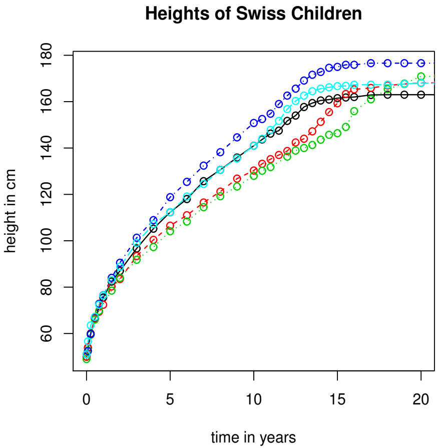
Cody Carroll

University of California, Davis

June 1, 2018

What is functional data?

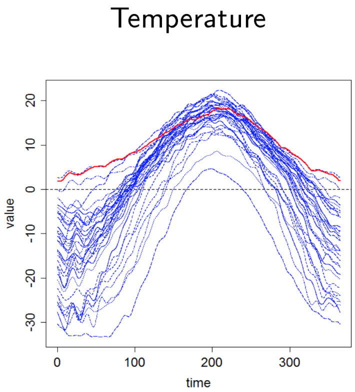
What are the striking features of the following sets of data?



Sample of human growth curves recorded in the Zürich Longitudinal Study.

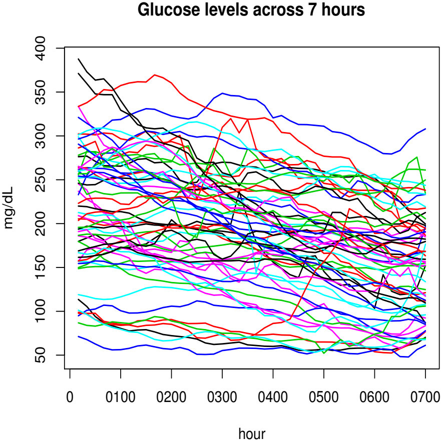
Examples

Temperatures (in $^\circ$C) over the course of a year measured at thirty-five weather stations across Canada.



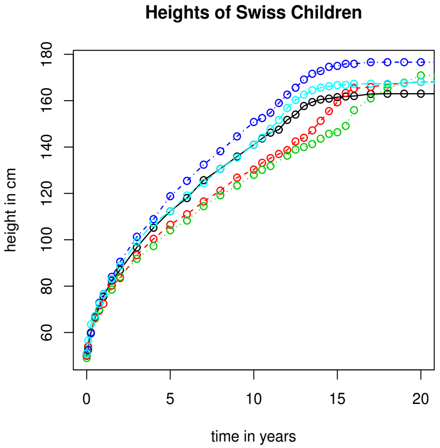
Examples

An individual's blood glucose levels measured over 7-hour window for 59 days.



What is functional data?

What are the striking features of the following data?



Features:

- similar shape
- repeated observations
- frequency
- smoothness

What is functional data?

- Functional Data: A sample of random functions, with one function per subject.
  - These functions can be curves (1D), images (2D or 3D), or have an higher number of components.
- Characteristics of functional data:
  (i) The "unit" of functional data is a function, $X_i(t)$.
  (ii) Functions are (in theory) $\infty-$dimensional data. There are an infinite number of time points to observe the function on.

Functional Data Basics

In theory, data $X_1, \ldots, X_n$ are i.i.d. copies of a random function $X(t)$.
In practice, the observed data for subject i is actually:

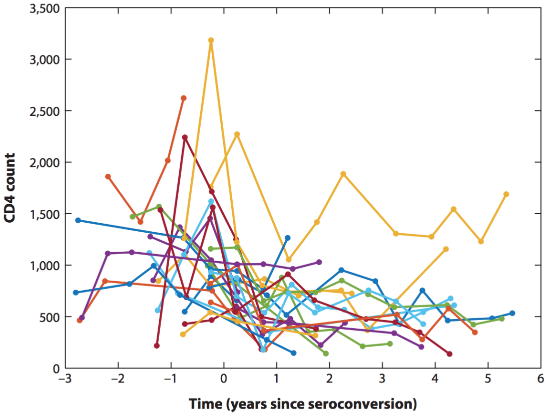
$$Y_{ij} = X_i(t_{ij}) + e_{ij}, \quad j = 1, \ldots, n_i,$$

where $X_i(t)$ is a smooth random function and e_{ij} are independent realizations of random noise for all i, j.

- Regular functional data - All subjects are measured at the same time grid, $t_1, \ldots, t_m \implies$ (often high-dim.) multivariate data.
- Irregular functional data - The measurement schedule for subject i is $t_{i1}, \ldots, t_{in_i} \implies$ longitudinal data.

Longitudinal Example

Irregularly recorded CD4 T lymphocyte count data for 25 HIV+ patients.

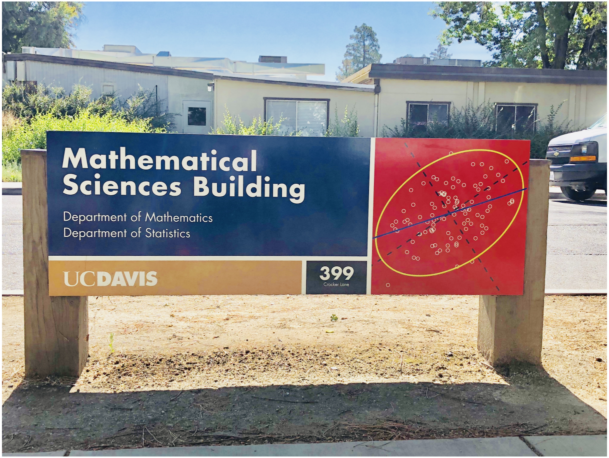


What are we interested in?

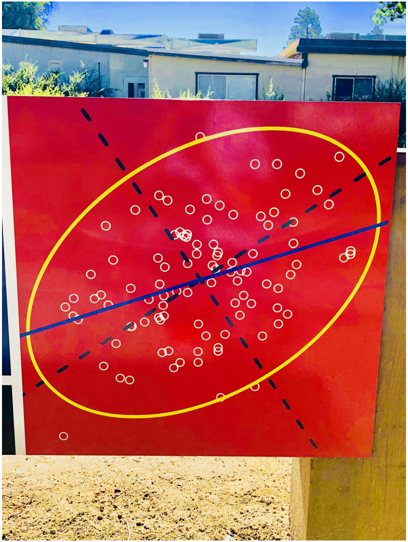
- Estimating the sample functions from finitely many observed points
- Identifying important sources of patterns/variation among the data
- Other more complicated questions

We need a statistical tool to accomplish these tasks. Let's take some inspiration from outside...

Outside this building



Principal Components Analysis (PCA)

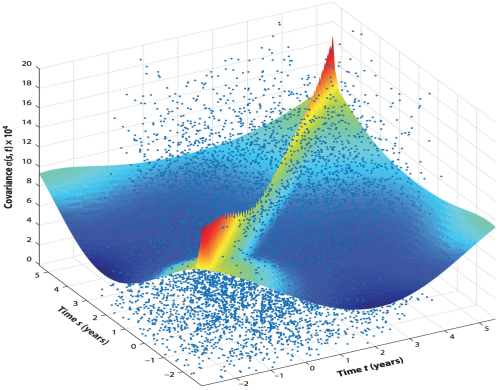


PCA

- PCA takes multivariate data (in $\mathbb{R}^p$) and finds the orthogonal "directions" (vectors) which best explain the variation in the data
- Each observation gets a "score" per direction, indicating how far in that direction the observation lies
- Often we only need 2 or 3 "directions" and their scores for an observation to recreate the data accurately, even when p is large
- Optional: mathematically, these directions are the eigenvectors of the sample covariance matrix of the data

Extend PCA to FPCA

- We want to find the "directions" which explain variation for functional data now
- In functional data, we have covariance surfaces instead of covariance matrices

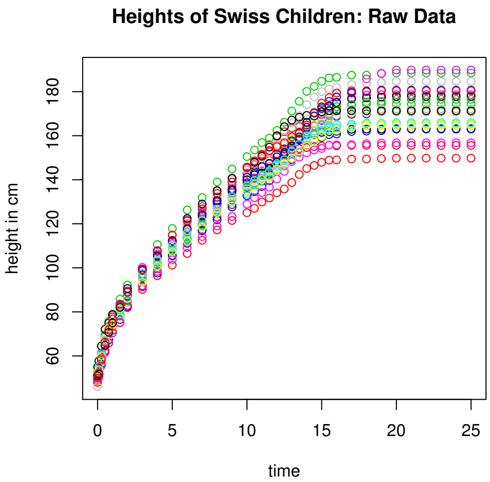


FPCA

- With some math we can obtain the best “directions” from the sample covariance surface which are called eigenfunctions
- Each observation will receive one score per eigenfunction, like before
- Often we only need 2 or 3 eigenfunctions to explain most of the variation, like in PCA
- Also, these eigenfunctions often have physical interpretations

FPCA Example

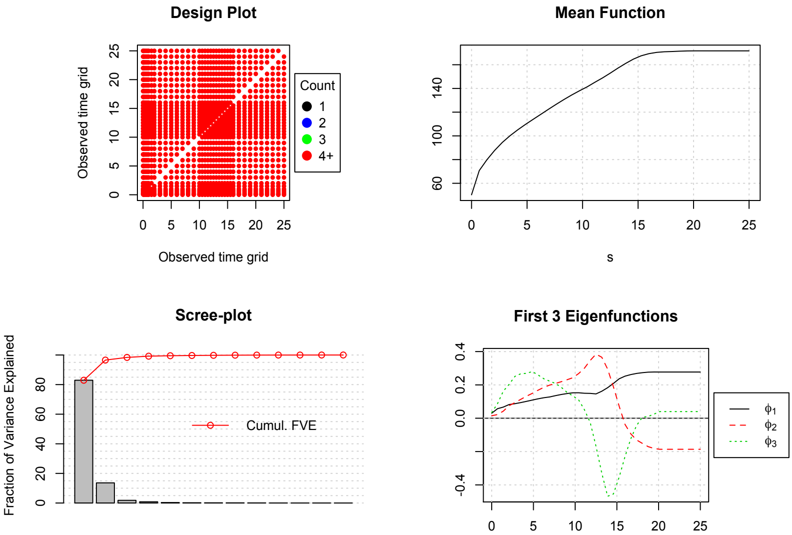
25 human growth curves recorded in the Zürich Longitudinal Study.



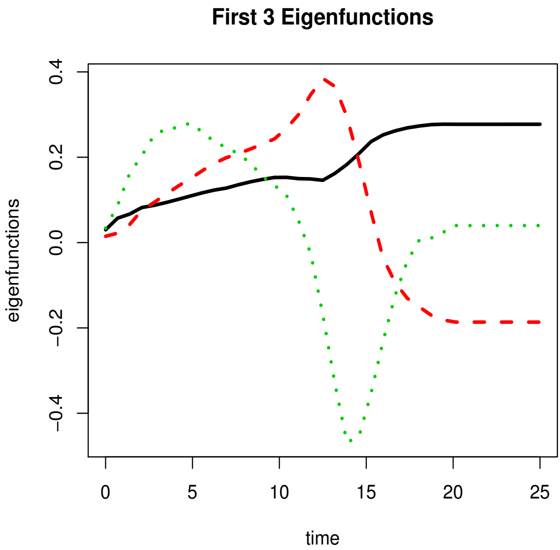
some math later...

FPCA Example

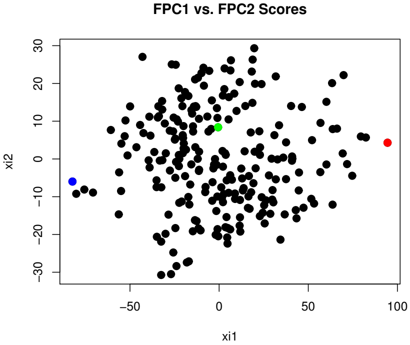
Using the `FPCA()` function in the `fdapace` package:

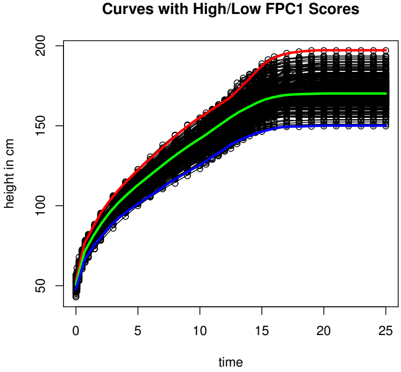


FPCA Example

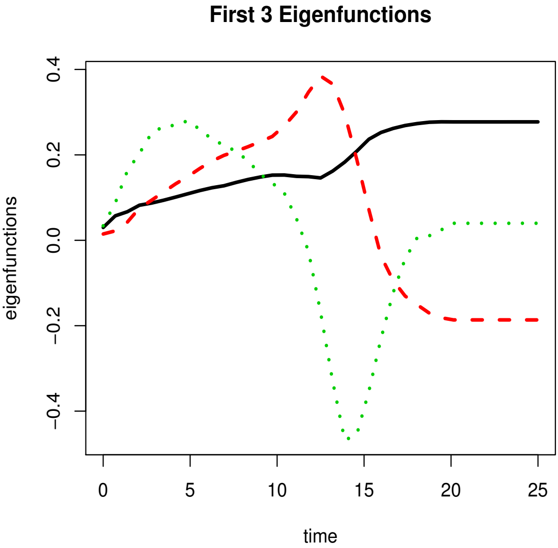


FPCA Example

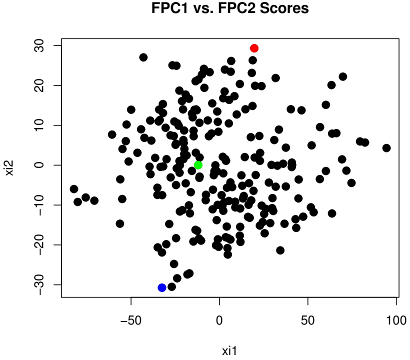


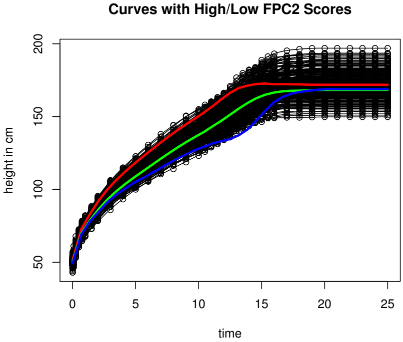


FPCA Example



FPCA Example





More complicated questions in FDA

- Functional regression:
  - function-on-function
  - function-on-scalar
  - scalar-on-function
- Questions of inference
- Functional registration
- Density estimation

References I

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., Molinari, L., and Prader, A. (1984a),
Velocity and Acceleration of Height Growth Using Kernel Estimation,
Annals of Human Biology, 11, 397411.

Hooker, G.
Functional Data Analysis: A Short Course.

Lu, X. and Marron, J. S. (2013),
Principal nested spheres for time warped functional data analysis,
Preprint.

Wang, J.-L. and Chiou, J.-M. and Müller, H.-G. (2016),
Review of Functional Data Analysis,
arXiv preprint.