

Functional Data Analysis

Jiguo Cao, PhD

Canada Research Chair in Data Science

Director of the Pacific Blue Cross Health Informatics Laboratory

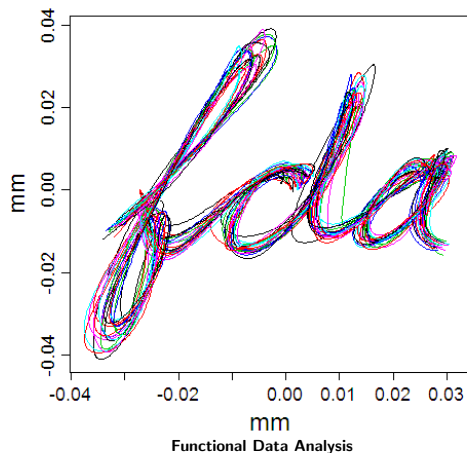
Associate Professor, Department of Statistics and Actuarial Science

Associate Faculty Member, School of Computing Science

Simon Fraser University, Vancouver, Canada

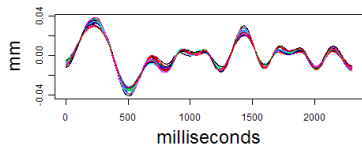
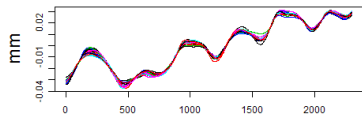
Classical Functional Data

Measures of position of nib of a pen writing "fda". 20 replications, measurements taken at 200 hertz.



What is Functional Data?

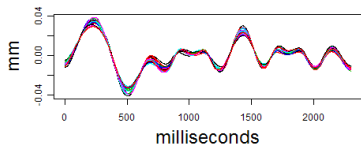
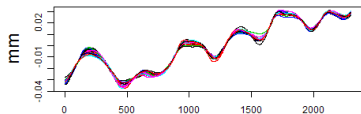
20 replications, 1401 observations within replications, 2 dimensions



What is Functional Data?

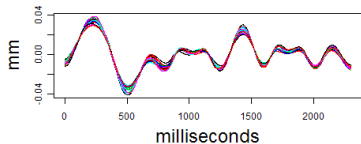
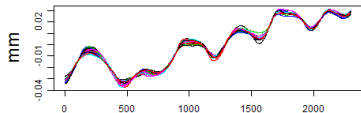
20 replications, 1401 observations within replications, 2 dimensions

Functional data is often complex:



What is Functional Data?

20 replications, 1401 observations within replications, 2 dimensions

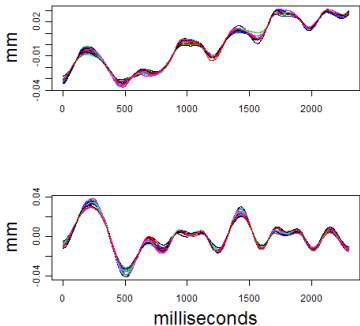


Functional data is often complex:

- often a large number of related quantities

What is Functional Data?

20 replications, 1401 observations within replications, 2 dimensions

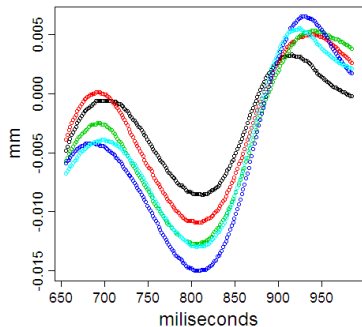


Functional data is often complex:

- often a large number of related quantities
- viewing each replication as a single observation can make the data easier to think about (once we have the right machinery)

What is Functional Data?

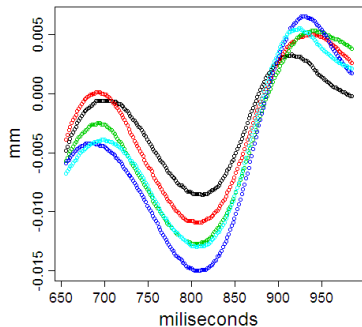
What are the most obvious features of these data?



What is Functional Data?

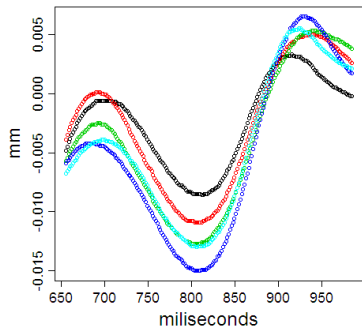
What are the most obvious features of these data?

- quantity



What is Functional Data?

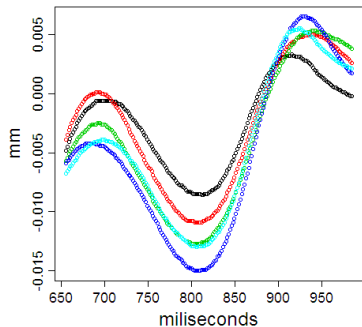
What are the most obvious features of these data?



- quantity
- frequency (resolution)

What is Functional Data?

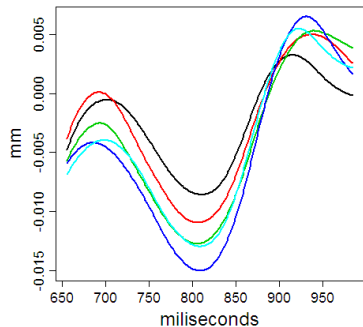
What are the most obvious features of these data?



- quantity
- frequency (resolution)
- similar trends

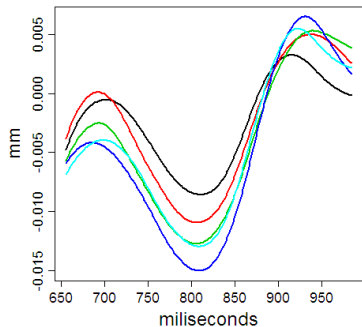
What is Functional Data?

Most important: **smoothness**



What is Functional Data?

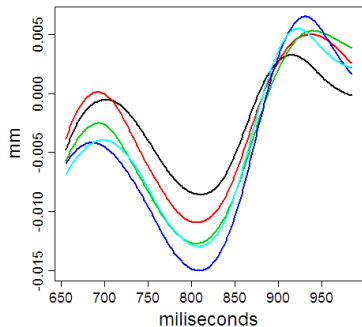
Most important: **smoothness**



These data describe (nearly) a process that changes smoothly, and continuously over time.

What is Functional Data?

Most important: **smoothness**

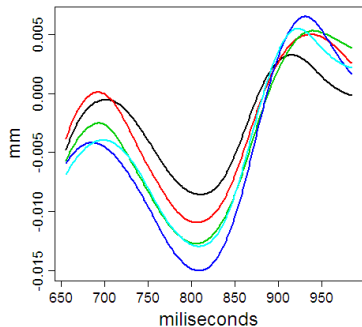


These data describe (nearly) a process that changes smoothly, and continuously over time.

Functional Data Analysis = Analysis of data that are functions.

What is Functional Data?

Most important: **smoothness**



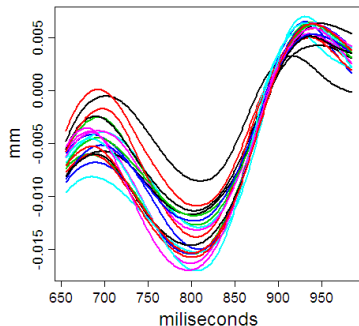
These data describe (nearly) a process that changes smoothly, and continuously over time.

Functional Data Analysis = Analysis of data that are functions.

Domain is usually time, but can be anything: space, energy ...

What is Functional Data?

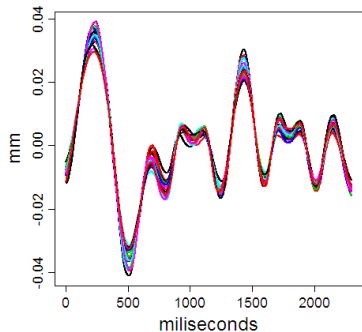
20 replications



Functional data analysis involves repeated measures of the same process.

What is Functional Data?

20 replications, 1401 observations within replications



Functional data is often complicated:

- not easily described by mathematical formulae
- variation between replications even harder to describe

Characteristics

- Data are measurements of smooth processes over time
- We usually do not want to make parametric assumptions about those processes.
- Often have multiple measurements of the same process
- We are interested in describing the variation of processes.
- Frequently, collected data have high resolution and low noise.
- Can be applied to any estimate of a smooth process.

About Functional Data Analysis

1. FDA is New

- First named in Dalzell & Ramsay, 1991
- Relatively little penetration into applied fields (= easy publication)
- Several competing methodologies (we focus on one)
- Limited public software/resources
- data analysis rather than inference

2. Functional Data is Complex

- Requires more thought/judgement than a t -test
- data needs pre-processing
- parametric inference is rarely available/appropriate

What will this course offer?

Audience: application areas with functional data

- Focus:**
- What can Functional Data Analysis do?
 - How do I make it happen?

Software: packages in R, Matlab

Goals: Enabling you to

- Understand and interpret the result of FDA applied to real data
- Use existing FDA libraries to analyze functional data
- Evaluate its usefulness/correctness
- Extend the methods in existing software if you need to

Not Covered: reproducing-kernel Hilbert spaces, asymptotics, theorems...

Pre-requisites and Recommendations

Pre-requisites: multiple linear regression

Useful: Life will be easier if you do not need to learn some of the following:

- R/Matlab or other programming experience
- Calculus
- Matrix algebra
- Multivariate statistics
- Computational statistics

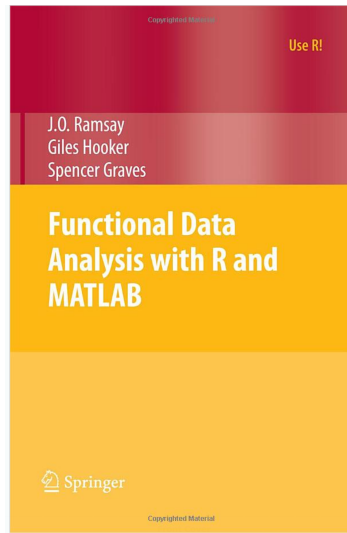
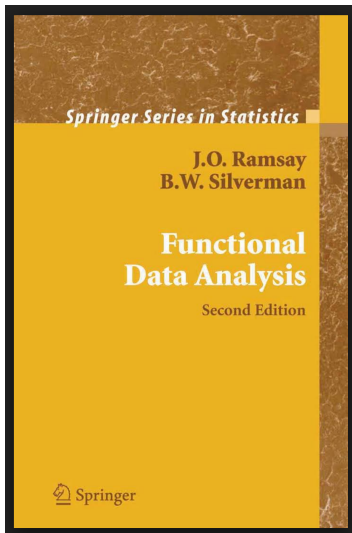
Any necessary material will be covered in class, but will be out of context.

Resources

Textbook: Ramsay and Silverman, 2005, Functional Data Analysis, Springer.

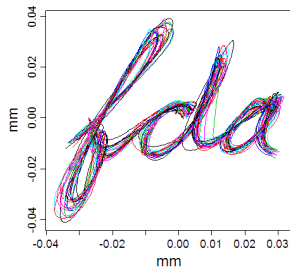
- Books:**
- Ramsay and Silverman, 2002, Applied Functional Data Analysis, Springer.
 - Ramsay, Hooker and Graves, 2009, Functional Data Analysis with R and MATLAB, Springer

- Online:**
- <https://github.com/caojiguo/FDAcourse2019/>
All lecture slides, exercises, computer codes, etc will be posted here.
 - <https://www.youtube.com/channel/UC1Wh20PhCEOnrEkk58WweFA>
Recorded lectures will be posted here.
 - <http://www.functionaldata.org> for FDA
 - <http://www.r-project.org> a general site for R



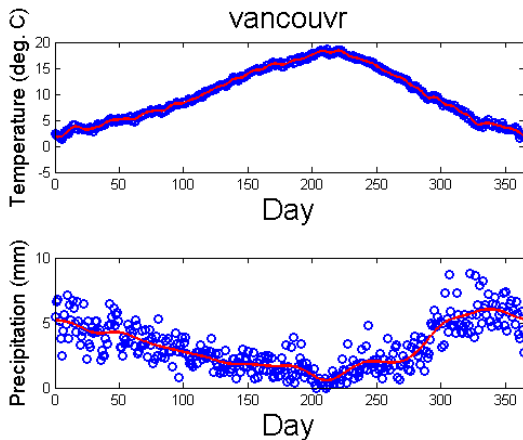
Back to "What is Functional Data"

Or What isn't Functional Data?



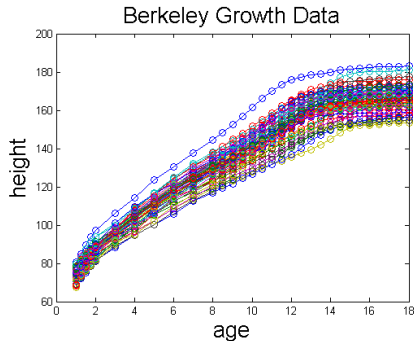
Do my data need to look this good?

Data may be measured more noisily



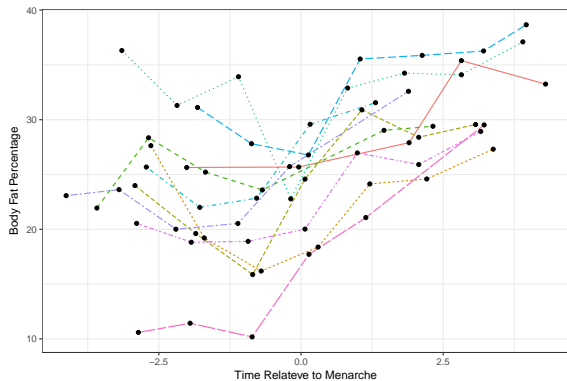
We need to find the smooth process under the data.

Data may be measured more sparsely



- Data are low noise but low-resolution
- Measured at unequal intervals
- We know that the curves must always increase

MIT Growth and Development Study



- 162 girls
- 6.4 measurements on average (sparse)
- irregular

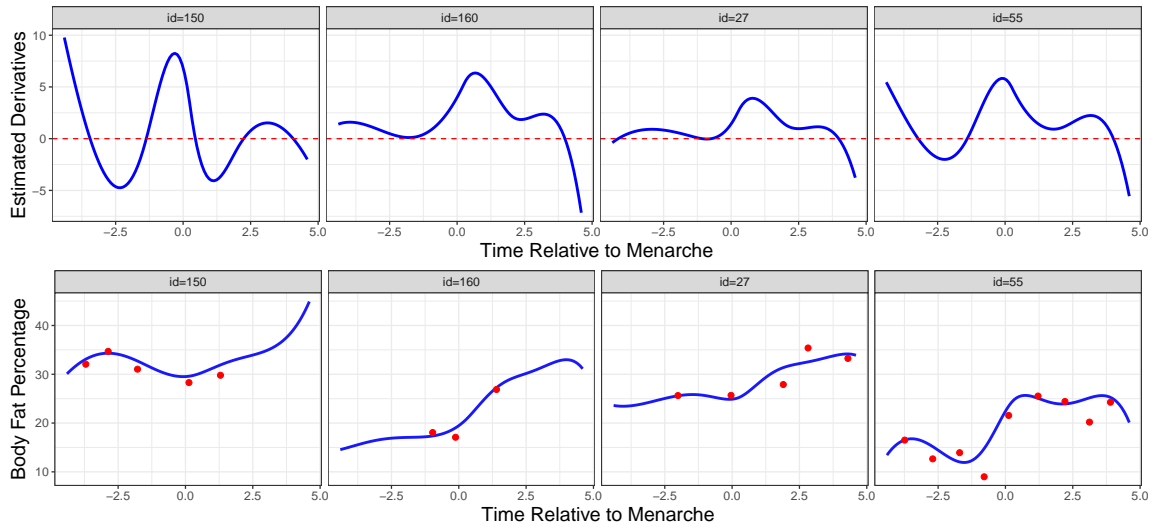
Our goal

- Estimating the underlying functions

Our goal

- Estimating the underlying functions
- Estimating the derivative functions

For instance...



Necessities for Functional Data

- must believably derive from a smooth process
- process should not be easily parameterizable (should not be able to write down a formula)
- enough data to resolve the essential features of the process (peaks, zero-crossings, speed... will depend on application)
- some repetition in the process
- do not need equally-spaced or perfect measurements

Common Sources

medical monitoring: EEG, ECG, fMRI, blood pressure ...

medical tests: HIV antibodies, flu screens...

biology: animal behavior (whale songs, fly egg-laying...)

environmental monitoring: weather, pollution, solar radiation, traffic ...

optotrack experiments: psychology/physiology

economics/marketing: macro-trends, futures markets

web data: e-bay auction prices, google trends

Essential Questions

Or what can FDA do for me?

- How do we go from discrete to functional data?
- How do we describe random variation in functional data?
- How do we decide if groups of functional data are different?
- How do we relate functional data to other data? To other functional data?
- What is special about functional data?
 - Aligning functions (registration)
 - Use of rates of change (dynamics)

Approximate Class Agenda

1. From data to functional data

- Basis expansions and smoothing
- The `fda` library
- Positive and monotone smoothing

2. Exploring Functional Data

- Means, variances, covariances
- Functional PCA

3. Functional Linear Models