M7777 Applied Functional Data Analysis
1. Introduction

Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Course Outline and Requirements

Outline

1. Introduction
2. Basis Systems
3. Basis Smoothing
4. Smoothing Penalties
5. Constrained Smoothing
6. Exploratory Data Analysis, FPCA
7. Scalar-on-function Regression
8. Functional Data Simulation
9. Function-on-scalar Regression
10. Function-on-function Regression
11. Registration
12. Sparse FDA

Course Outline and Requirements

Assumed Knowledge

This class will focus on the application of functional data analysis techniques to real-world problems and is not intended to be mathematically technical. However, we will make use of linear algebra and I assume a background in applied statistics on the level of M5120.

Computing Software

The course will be taught using the fda library in ℝ. I do not assume knowledge of ℝ, but some programming experience will be helpful. ℝ is freely available from `www.r-project.org`.
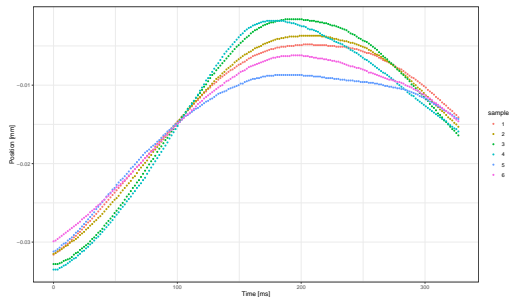
Assessment

- Attendance
- Homework
- Final Project

Students are expected to work individually on homework. The project may be undertaken in small groups.
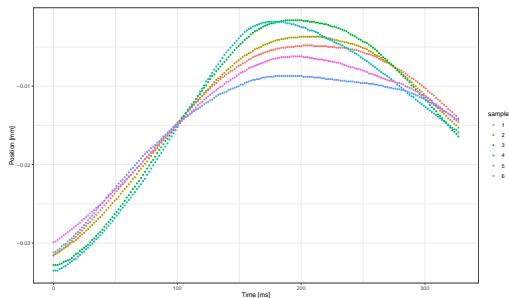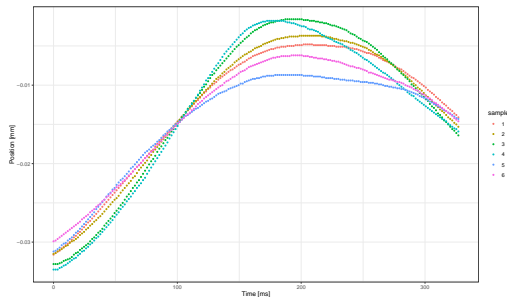
What are the most obvious features of these data?

- quantity

What are the most obvious features of these data?
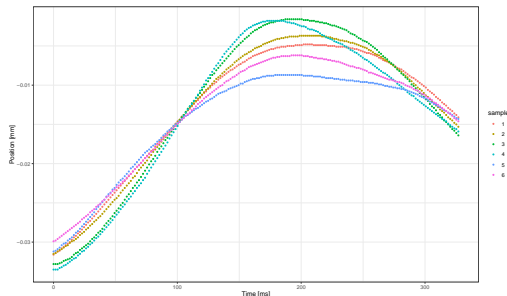

- quantity
- frequency (resolution)

What are the most obvious features of these data?

- quantity
- frequency (resolution)
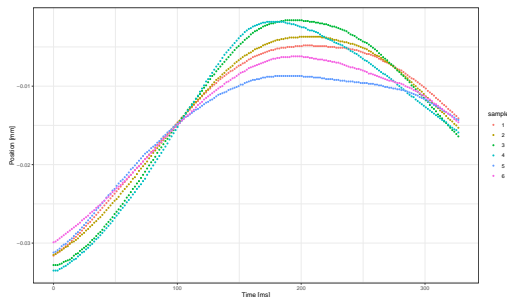- similar trends

What is Functional Data?

What are the most obvious features of these data?

- quantity
- frequency (resolution)
- similar trends
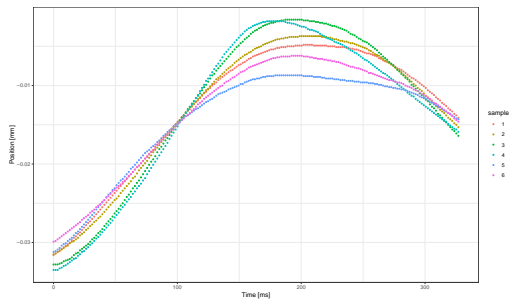- the same domain (not necessary)

What are the most obvious features of these data?

- quantity
- frequency (resolution)
- similar trends
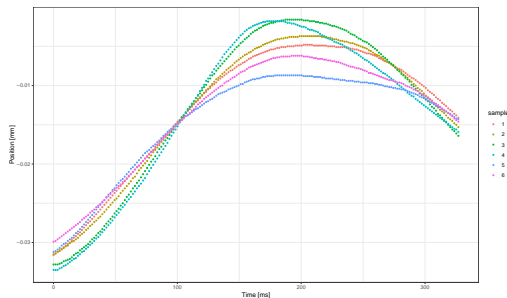- the same domain (not necessary)
- **smoothness**

6 replications, 200 observations within replications

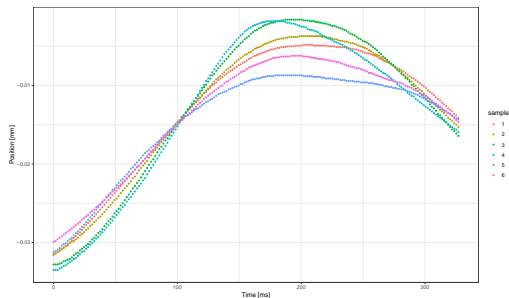- FDA involves repeated measures of the same process

What is Functional Data?

6 replications, 200 observations within replications

- FDA involves repeated measures of the same process
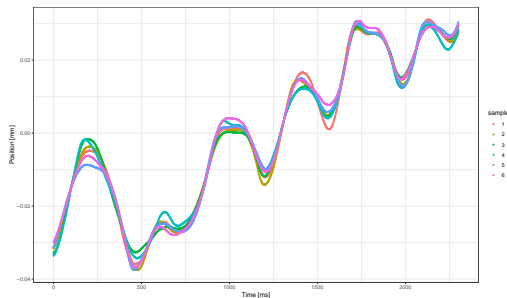- 1 observation = 1 function

6 replications, 200 observations within replications

- FDA involves repeated measures of the same process
- 1 observation = 1 function
- FDA = Analysis of data that are **functions**

6 replications, 1401 observations within replications
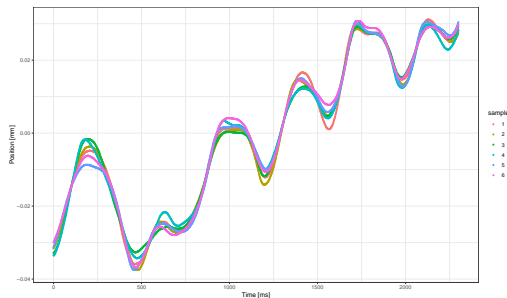

Functional data is often complicated:

- not easily described by mathematical formulae

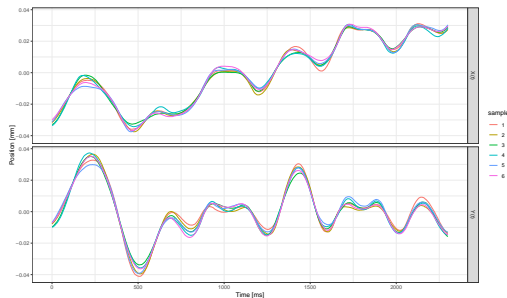6 replications, 1401 observations within replications

Functional data is often complicated:

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

What is Functional Data?

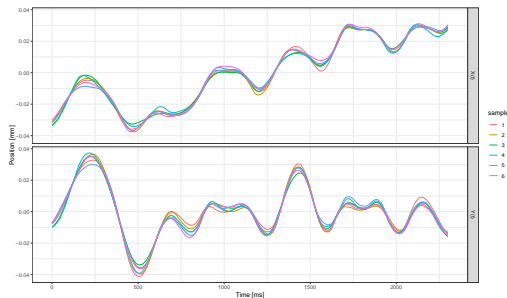6 replications, 1401 observations within replications, 2 dimensions

Functional data is often complex:

- often a large number of related quantities

6 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

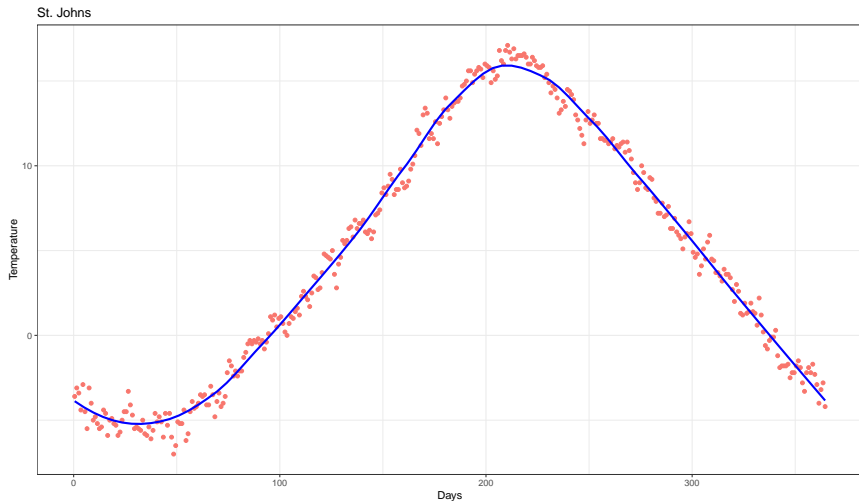What are these data? Let us plot one component against another!

What is Functional Data?

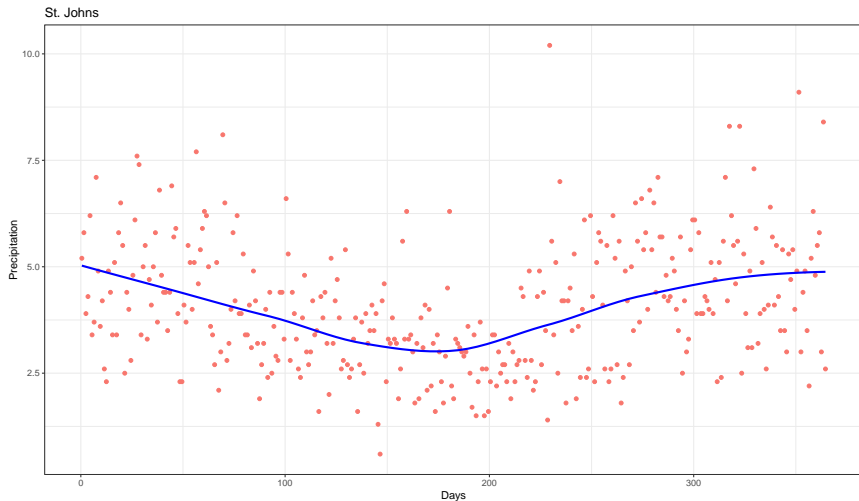Measures of position of nib of a pen writing "fda". 6 replications, measurements taken at 200 hertz.
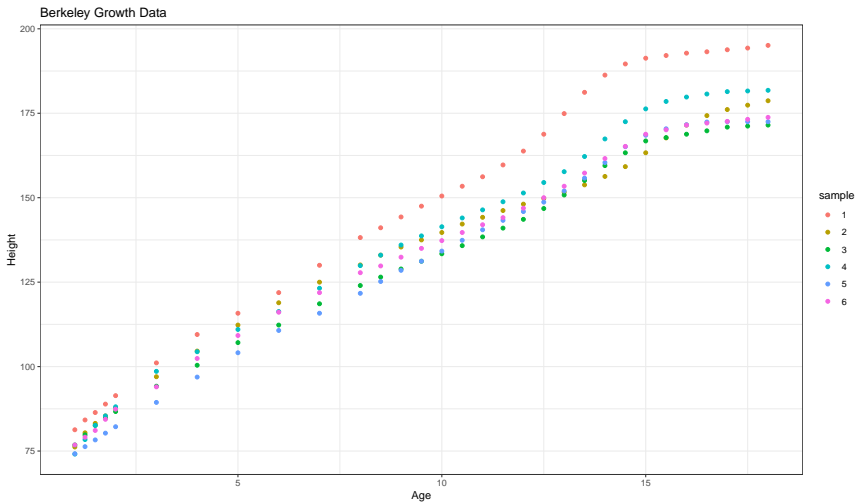
Data may be measured more noisily
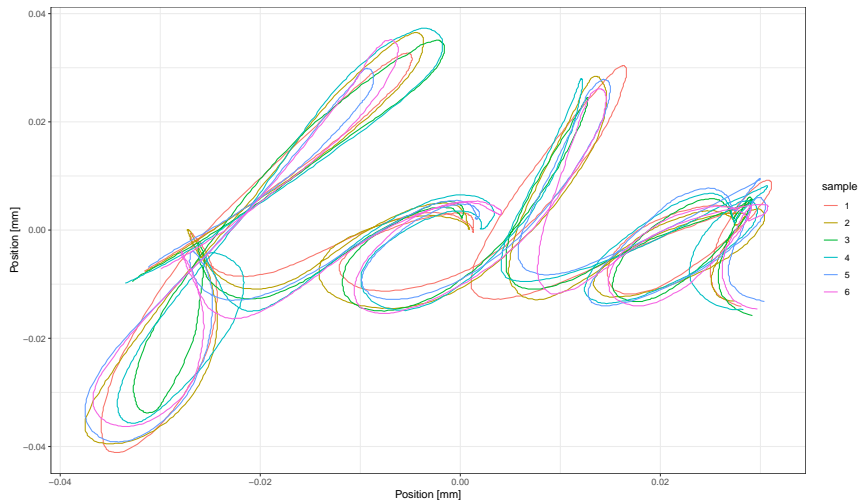
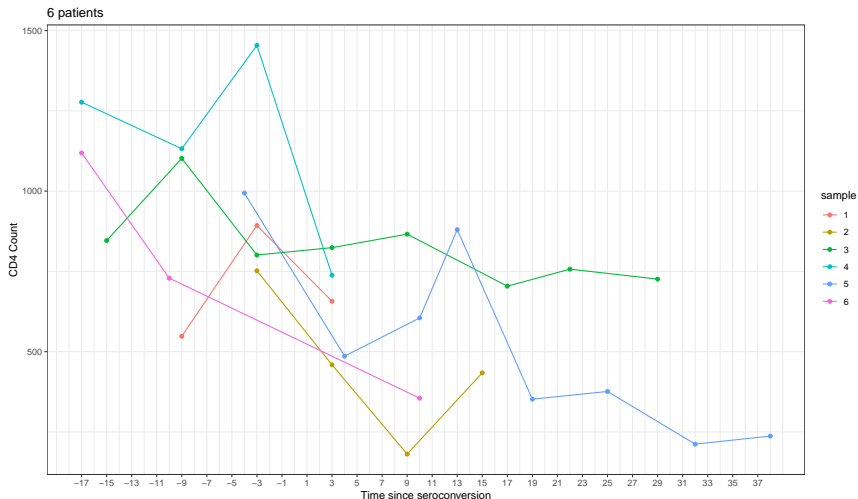# What is Functional Data?

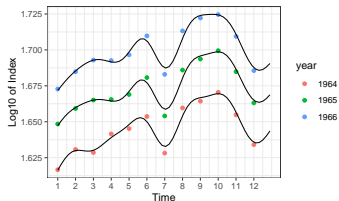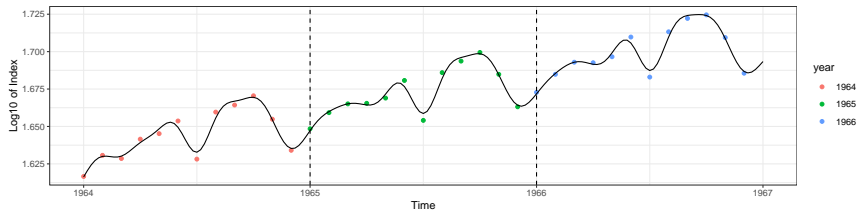## Data may be measured more noisily

Data may be measured more sparsely

What is Functional Data?

Longitudinal Data

We may not have repeated measurements

What is Functional Data?

From discrete to functional data – intuition

- The term **functional** in reference to observed data refers to the intrinsic structure of the data being functional; i.e. there is an underlying function that gives rise to the observed data.
- Advantages of representing the data as a smooth function:
  - allows evaluation at any time point
  - allows evaluation of rates of change of the underlying curve
  - allows registration to a common time-scale

Main idea in FDA: treat the observed data functions as single entities, rather than sequence of individual observations.

References

- Ferraty, F., Vieu, P., 2006. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

- Ramsay, J. O., Silverman, B. W., 2005. *Functional data analysis*, 2nd Edition. Springer, New York.

- Ramsay, J. O., Silverman, B. W., 2007. *Applied functional data analysis: methods and case studies*. Springer.

- Ramsay, J. O., Wickham, H., Graves, S., Hooker, G., 2019. fda: Functional Data Analysis. R package version 2.4.8. *https://CRAN.R-project.org/package=fda*

- Giles Hooker's course BTRY 6150 *http://faculty.bscb.cornell.edu/~hooker/*

- Kokoszka, P., Reimherr, M., 2017. *Introduction to functional data analysis*. Taylor & Francis Group.

Problems to solve

1 Install the fda package.

2 Berkeley Growth Data
- load the variable growth from the fda package
- plot the first 6 samples for boys and girls separately (see Figure 1)
- plot the first 6 samples for boys and girls into one plot (see Figure 2)

3 Canadian Weather Data
- load the variable CanadianWeather from the fda package
- plot temperatures measured in Edmonton, Halifax, Montreal and Ottawa (see Figure 3)
- plot precipitations observed in Edmonton, Halifax, Montreal and Ottawa (see Figure 4)
- plot temperatures for all the places in dependence on regions; into one plot and separately (see Figures 5 and 6)

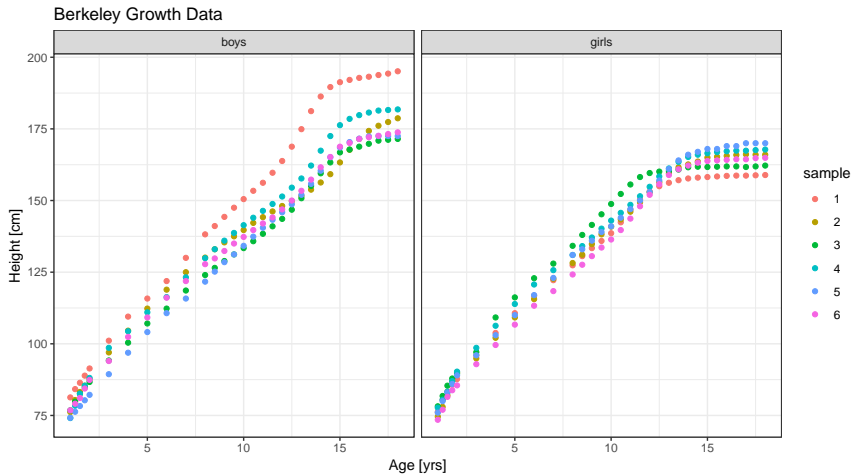**4** (optional) Plot other data from this presentation.

Figure 1.

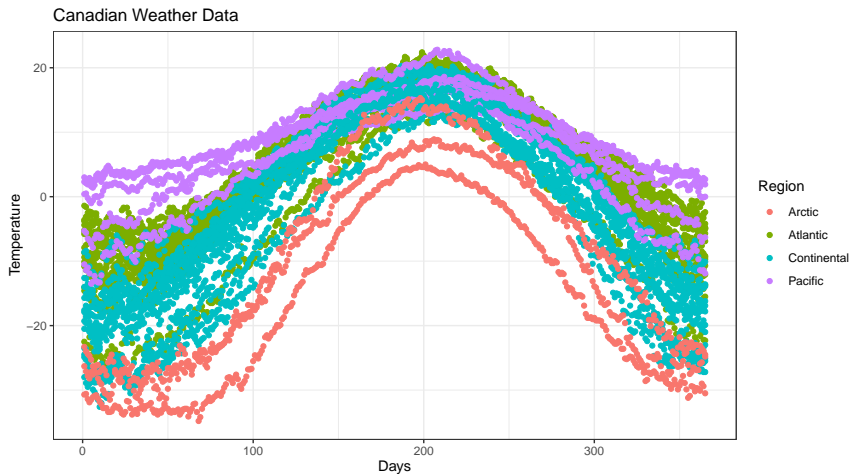Figure 2.

Figure 3.

Figure 4.

Figure 5.

Problems to solve

Figure 6.