

RStudio Energy Meetup

rstd.io/energy-meetup | March 23, 2022

Santiago Rodriguez | Consumers Energy



RStudio Enterprise
Community Meetup

RSTUDIO ENTERPRISE COMMUNITY MEETUP

Code of Conduct:

github.com/RStudioEnterpriseMeetup/codeofconduct

- Thank you all for making this a welcoming community
- Let's all be kind to one another
- We are dedicated to providing a harassment-free experience for everyone



Applications of Functional Data Analysis

Santiago Rodriguez
Data Scientist
Consumers Energy
Marketing



Agenda



- Introductions
- Functional Data Analysis (FDA)
- Load Profiles
- Derivatives
- Segmentation
- Resources

Introductions

About me



- Ecuador -> Florida -> Texas
- FSU + Texas A&M
- Data scientist in Marketing at Consumers Energy
 - Energy, aviation, automotive, contact centers
- Hobbies: Traveling, reading, fishing, and camping



TEXAS A&M
UNIVERSITY®

Image: https://en.wikipedia.org/wiki/Ecuador#/media/File:Flag_of_Ecuador.svg

Image: <https://brandguide.tamu.edu/visual-style/logos/downloads.html>

Image: <https://licensing.fsu.edu/brand-elements>

Image: https://en.wikipedia.org/wiki/Florida#/media/File:Seal_of_Florida.svg

Image: https://en.wikipedia.org/wiki/Texas#/media/File:Flag_of_Texas.svg

About Consumers Energy



6.8 Million

Michiganders count on us for service



8.7 Thousand

Workforce consisting of employees and contractors

Company was
Founded in
1886



87,097 Miles

of electrical
distribution lines



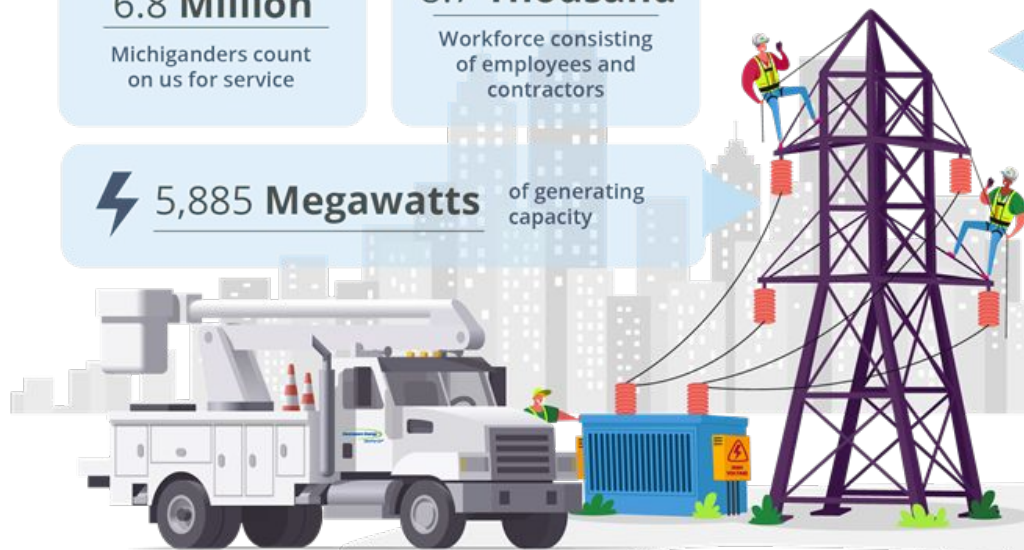
5,885 Megawatts

of generating
capacity



27,958 Miles

of natural gas
distribution pipeline



Consumers Energy

Count on Us®

The scope of today's meetup

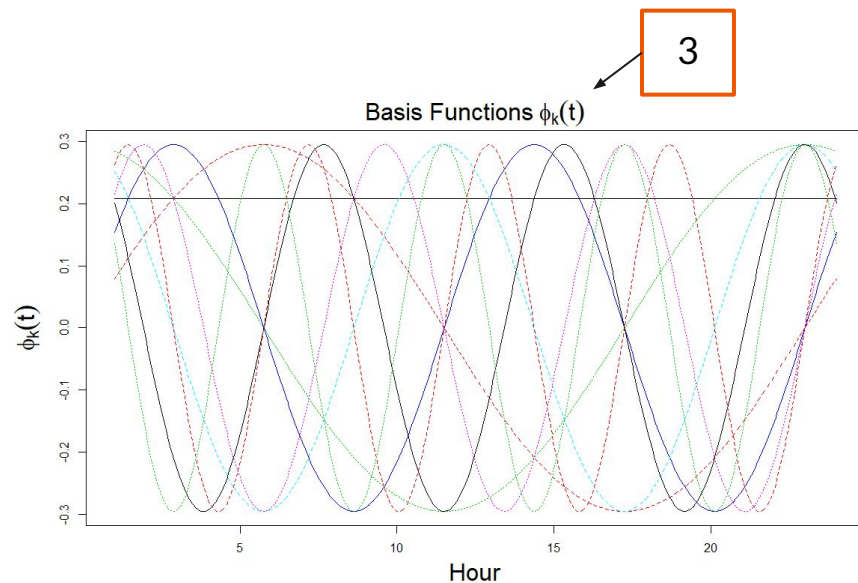
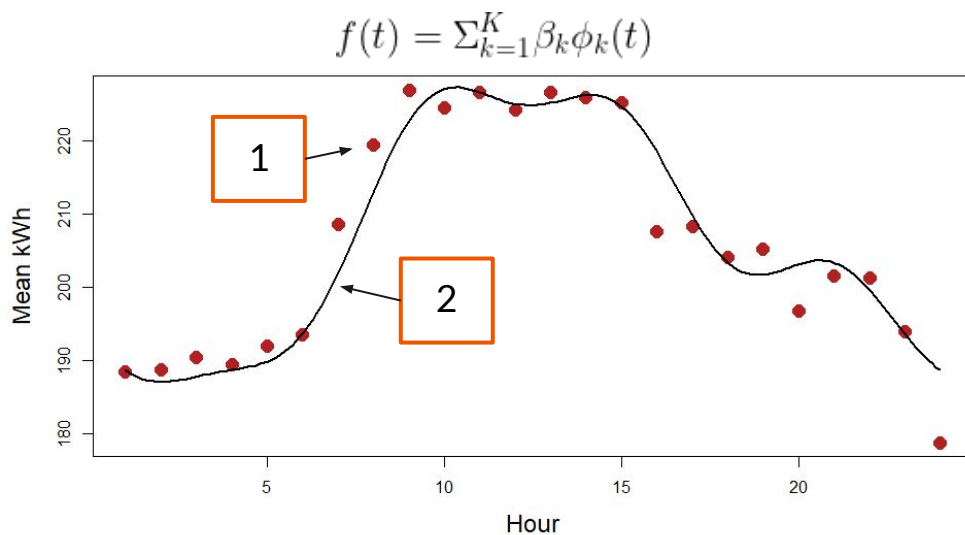
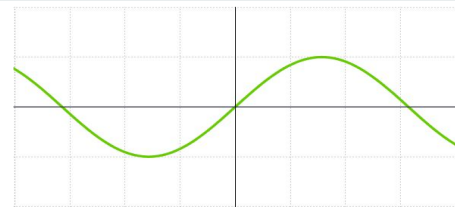
- The nature of today's talk will be **descriptive** and **non-technical**
- The goal today is to **introduce** functional data analysis and to demonstrate how it can **add value** to your work



Functional Data Analysis (FDA)

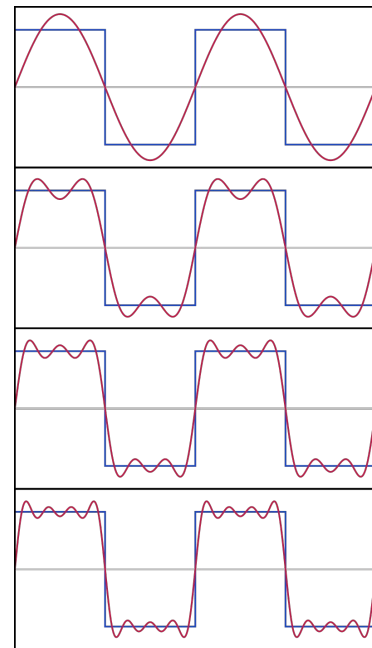
What is functional data analysis?

- The analysis of information on **curves** or functions
- Nonparametric, flexible regression technique
- Approximates a **curve** or function via a linear combination of basis functions




What can we do with these curves?

- Descriptive
(min, max, mean, median, variance, confidence intervals)
- Interpolation (connect the dots)
- Extrapolation (functional regression, GAMs)
- Clustering
- Can be used as part of traditional time series analysis



Why FDA?

- Another tool in your analytic 
- Flexible and tunable
 - Various ways to **fit** the curve: Fourier basis, B-splines, wavelet, etc.
 - Various ways to fit the **optimal** curve: least squares, penalized smoothing
- A unique feature of FDA is that the fitted curves are **differentiable**
 - Opens a new avenue of analysis that may tell us something that's difficult to see from the raw data itself



Pause



Have Questions?

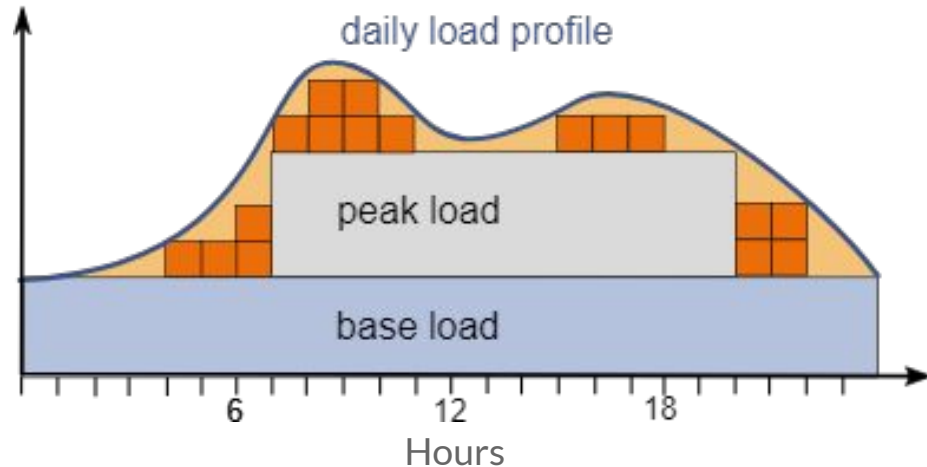
Load Profiles

What is a load profile?

“A load profile is a graph of the variation in the electrical load versus time”¹

Convenient way to
summarize the data

Helps us see
patterns more clearly



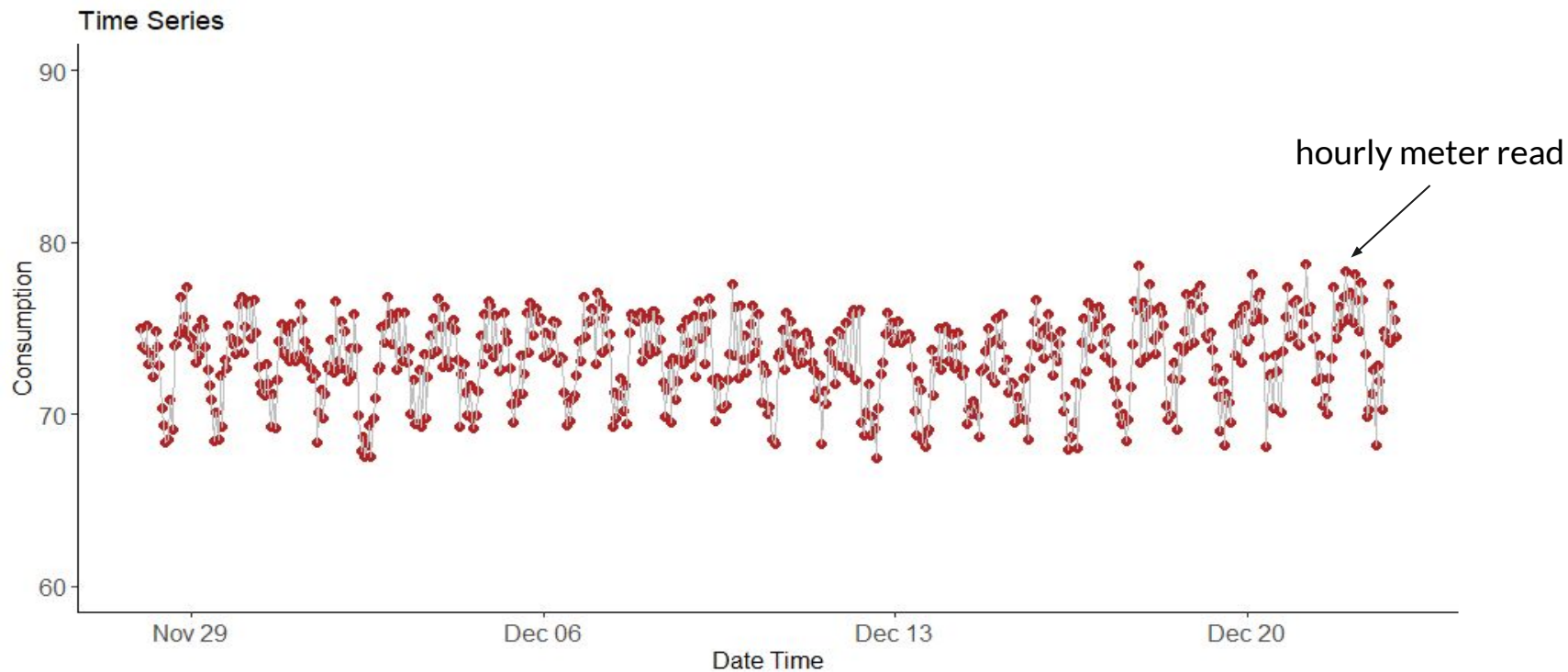
1. https://en.wikipedia.org/wiki/Load_profile
Image: https://commons.wikimedia.org/wiki/File:Hou710_ShapeRisk.svg

What can we do with a load profile?



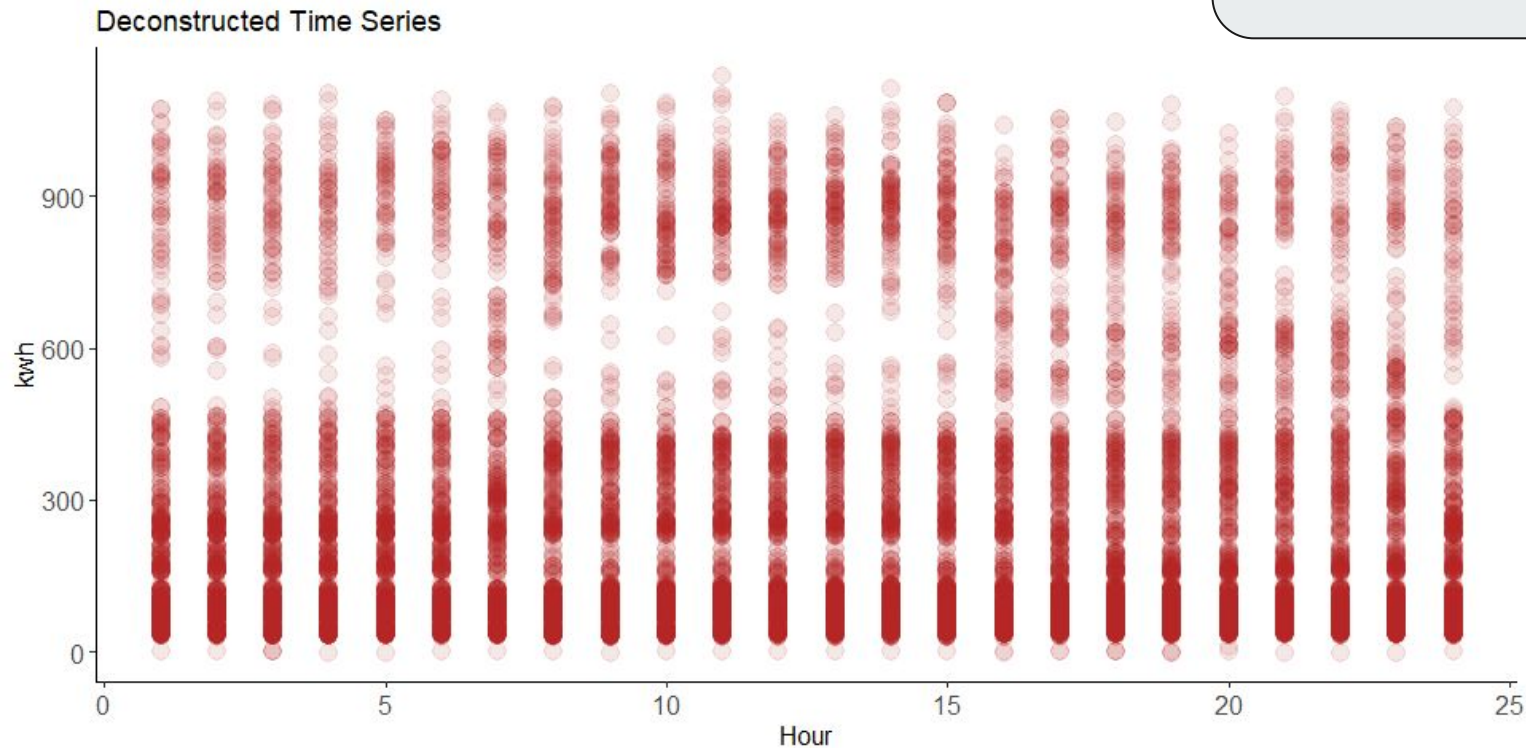
- Connect the dots using FDA
- Build confidence intervals to have a more complete picture
- Try different decompositions to create other non-conventional profiles and package the various profiles to capture different aspects of behavior

The raw data is a time series



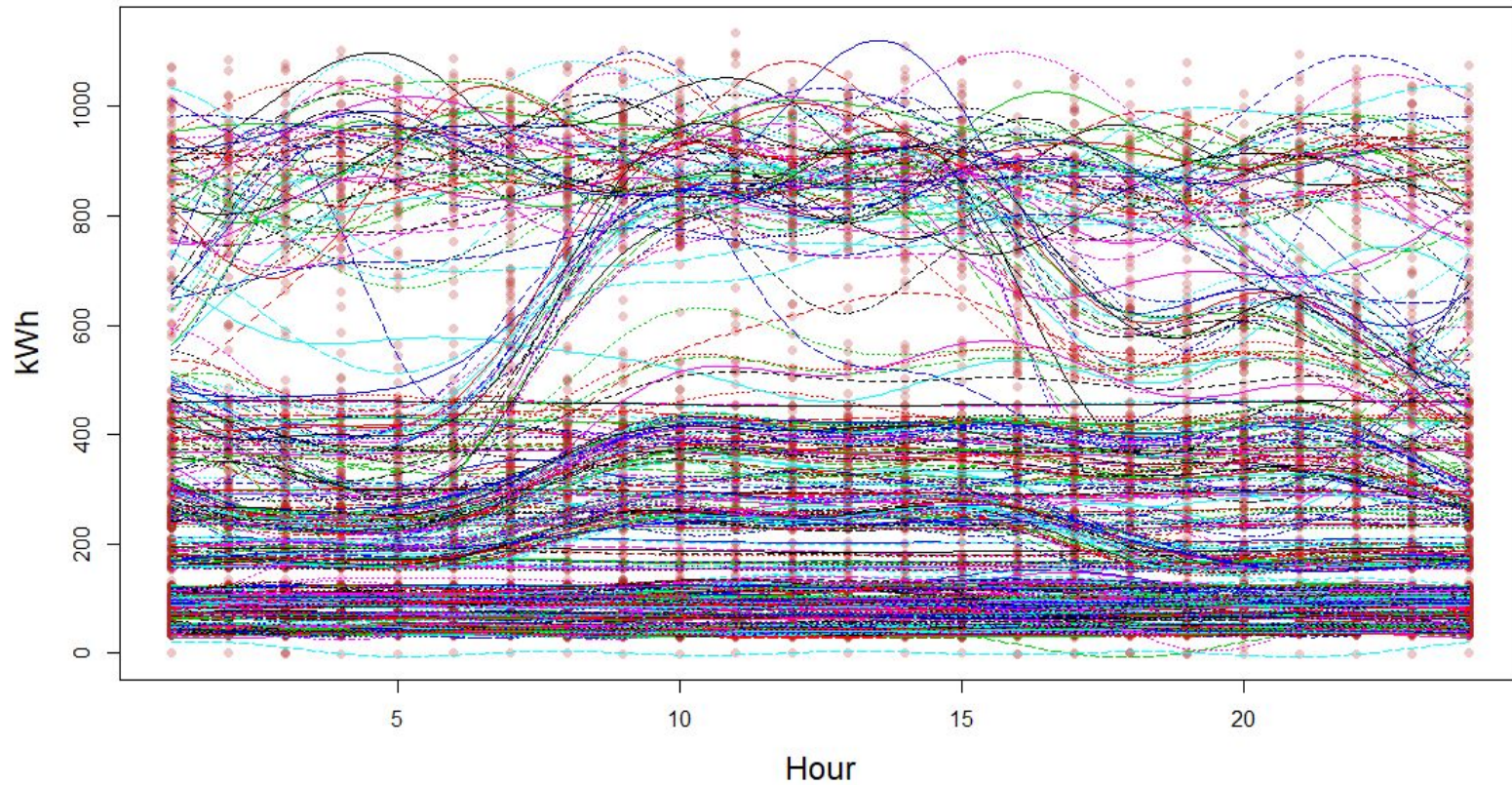
Step 1: Deconstruct the time series

Each dot represents consumption for a given hour and day

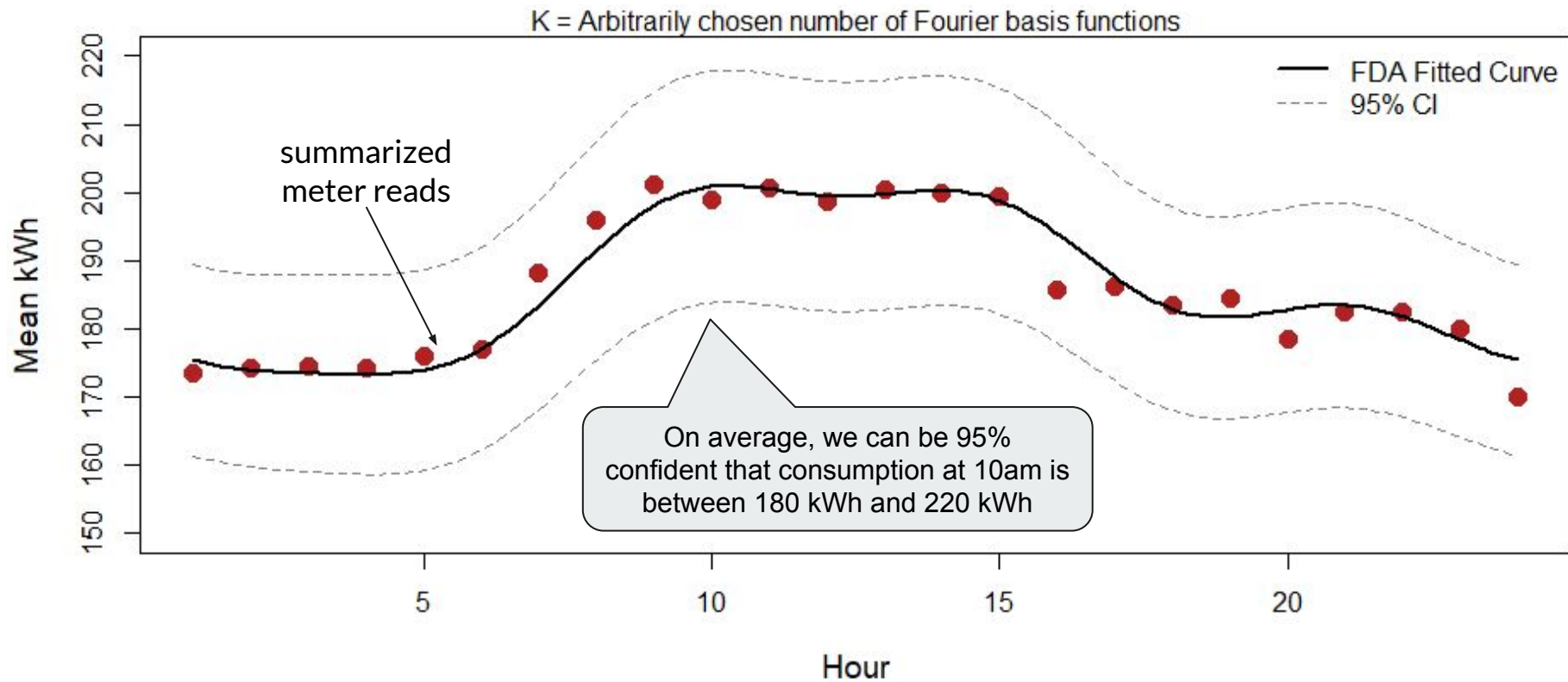


Step 2: Apply FDA

Used Fourier basis functions



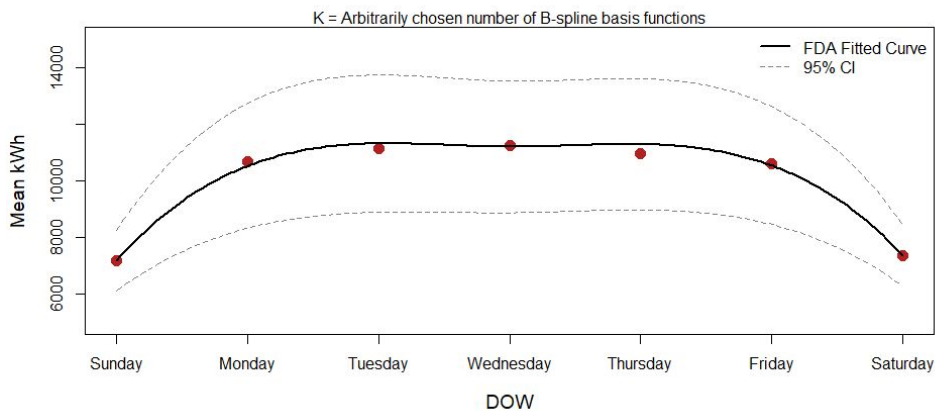
Step 3: Summarize



Other decompositions



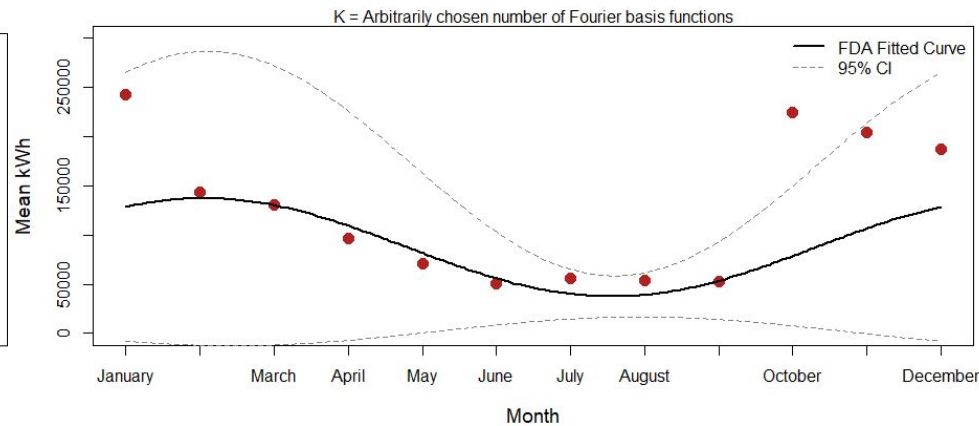
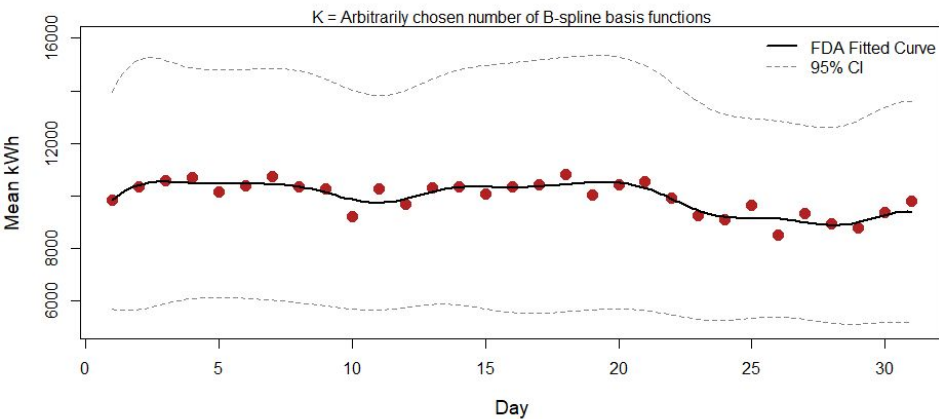
- Using the previously defined steps
we can construct other *less conventional* load profiles
- Such as:
 - Day of week (cubic splines)
 - Day of the month (cubic splines)
 - Month of the year (Fourier basis)



Looks like a frown lol



Idea: collect the load profiles and package them in a dashboard to easily see a customer's behavior



Pause

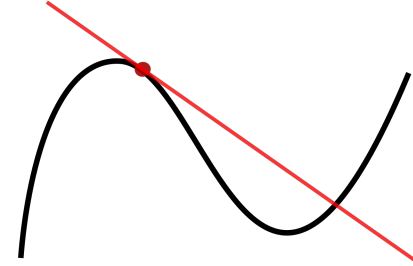


Have Questions?

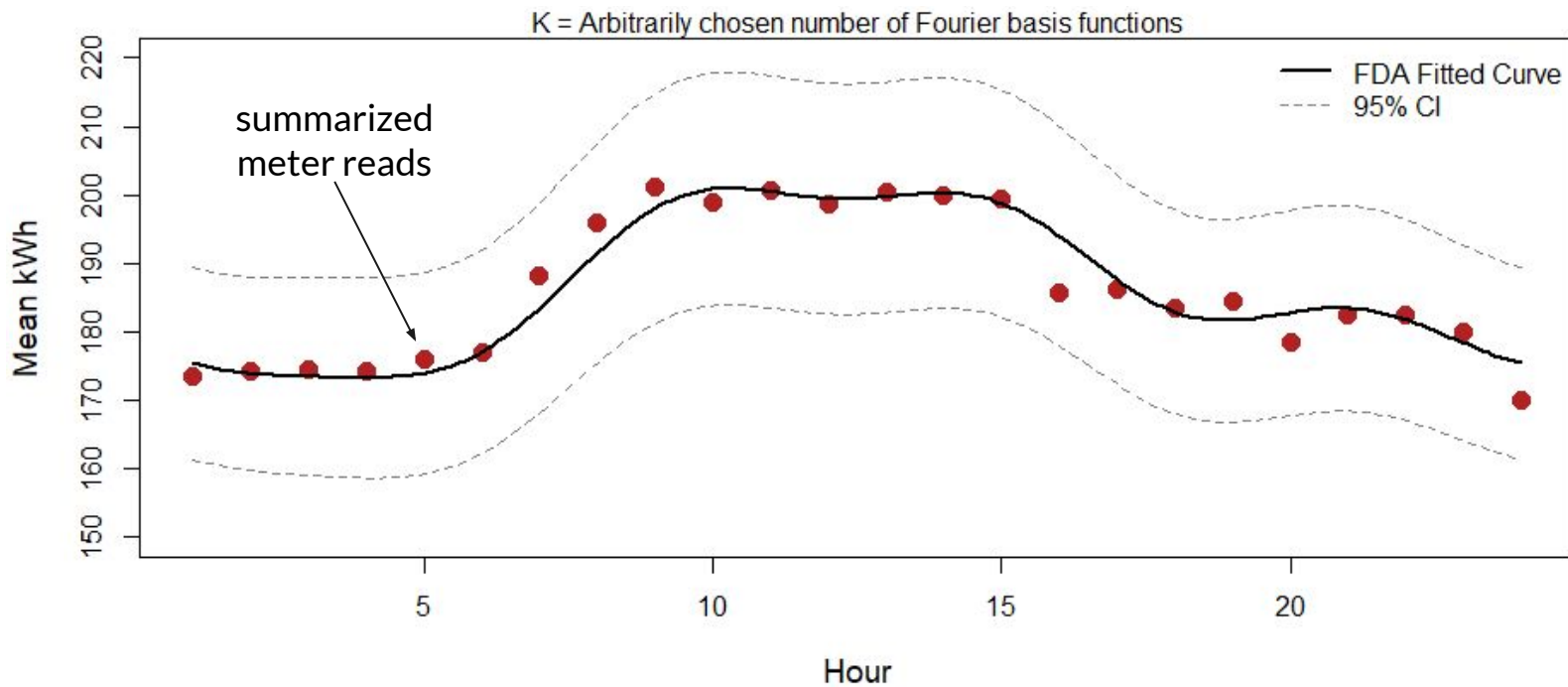
Derivatives

How are derivatives useful?

- Analyze rates of change
- The first derivative, velocity, tells us when a customer's consumption ramps up and down the fastest
 - Customer-centric work and personalization
 - Feature engineering
- Derivatives open a new realm to explore and analyze

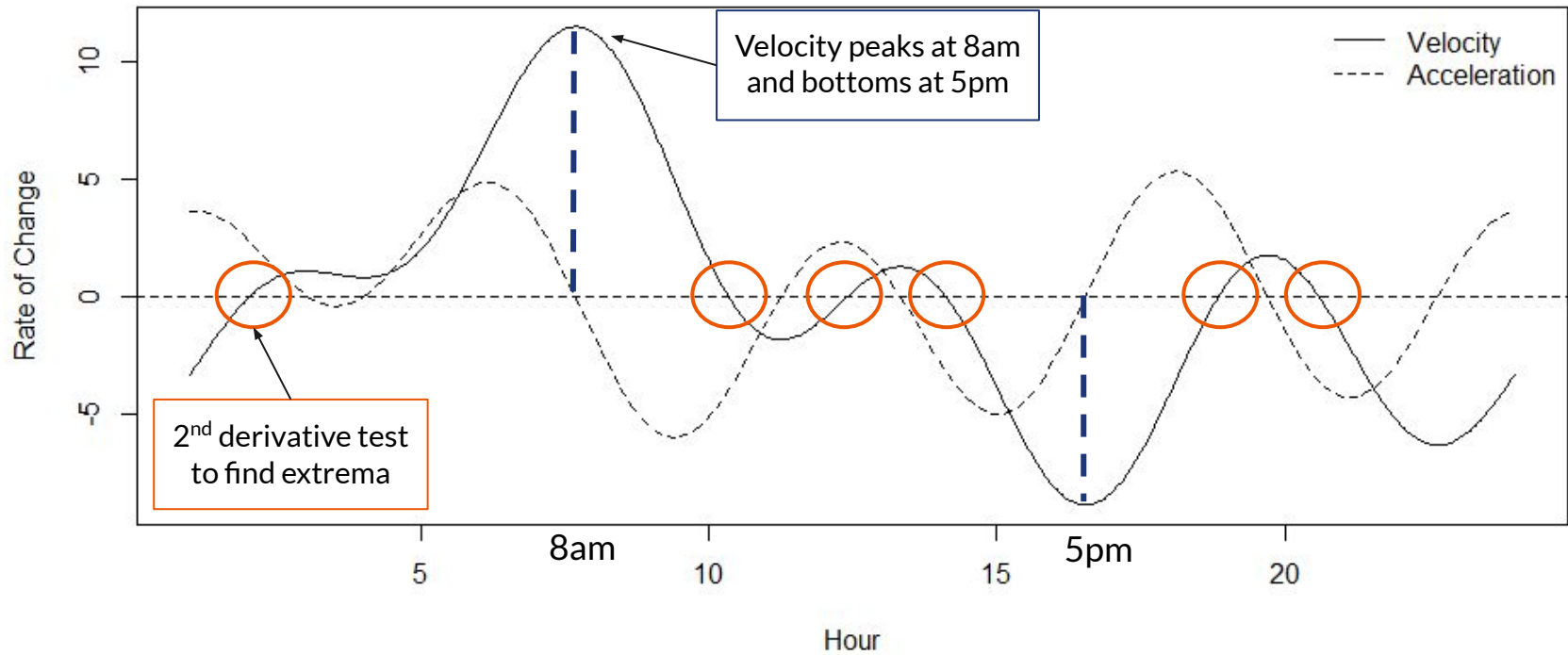


Daily load profile

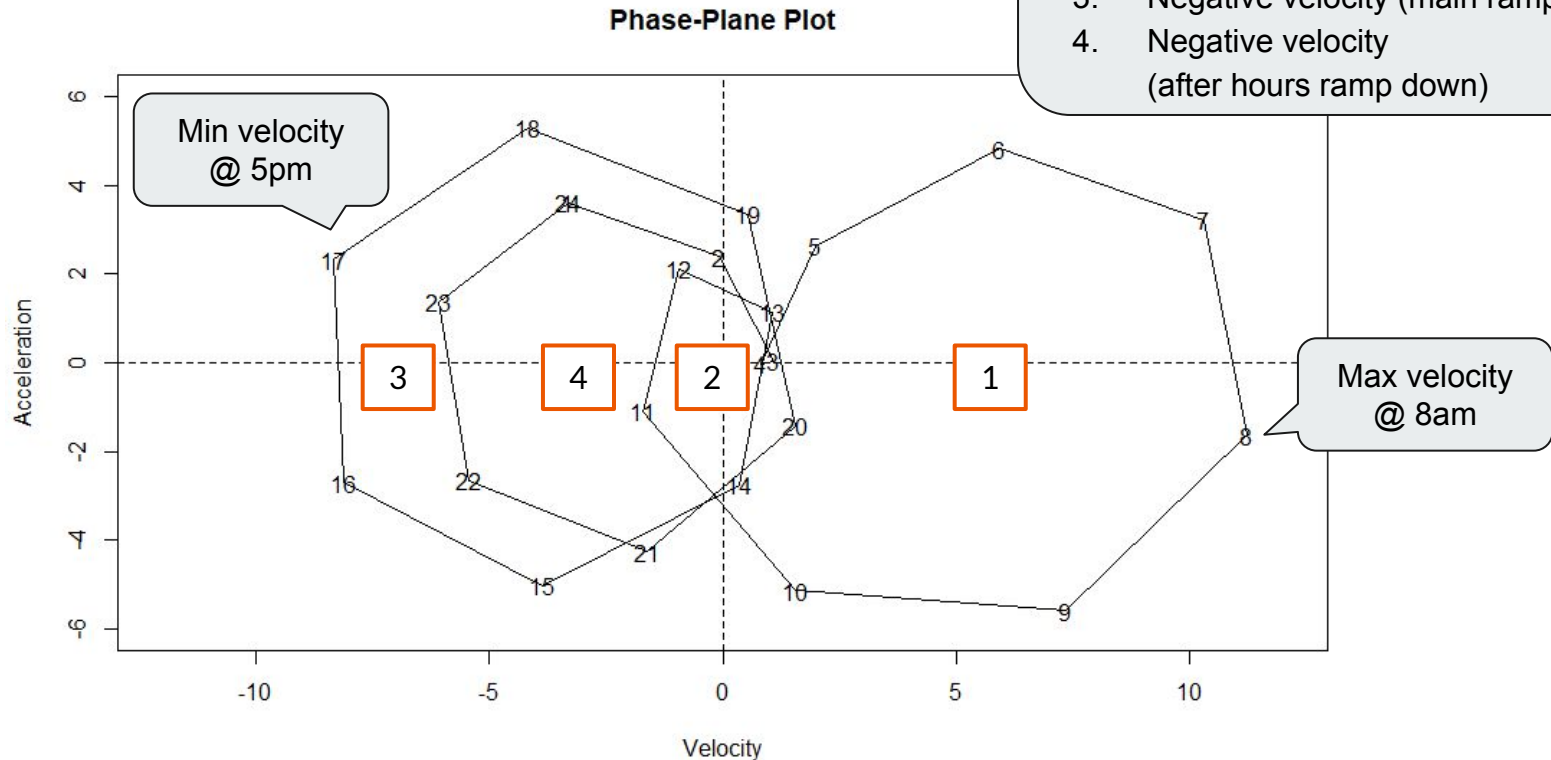


Using derivatives

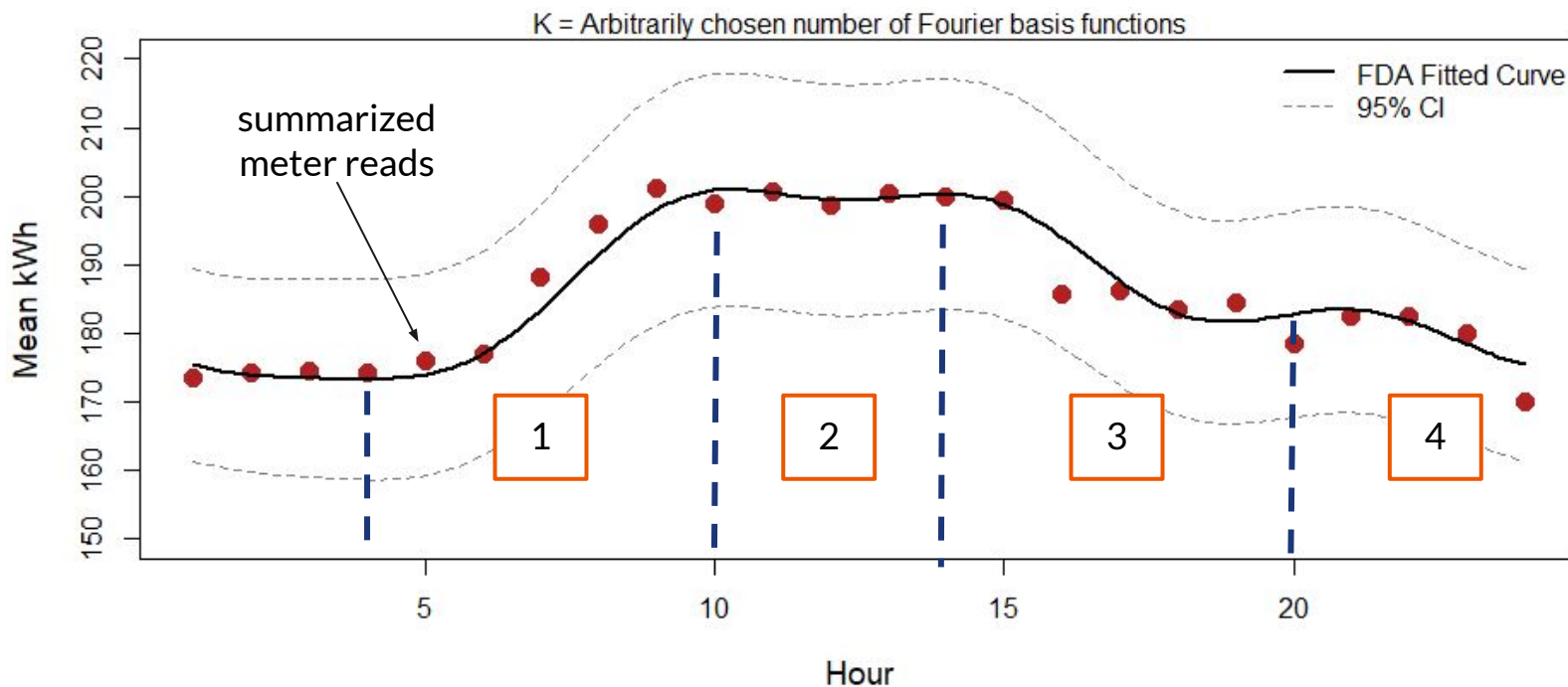
On average, we may consider on-peak for this customer to be between 8am - 5pm



Using derivatives, cont'd



Daily Load Profile



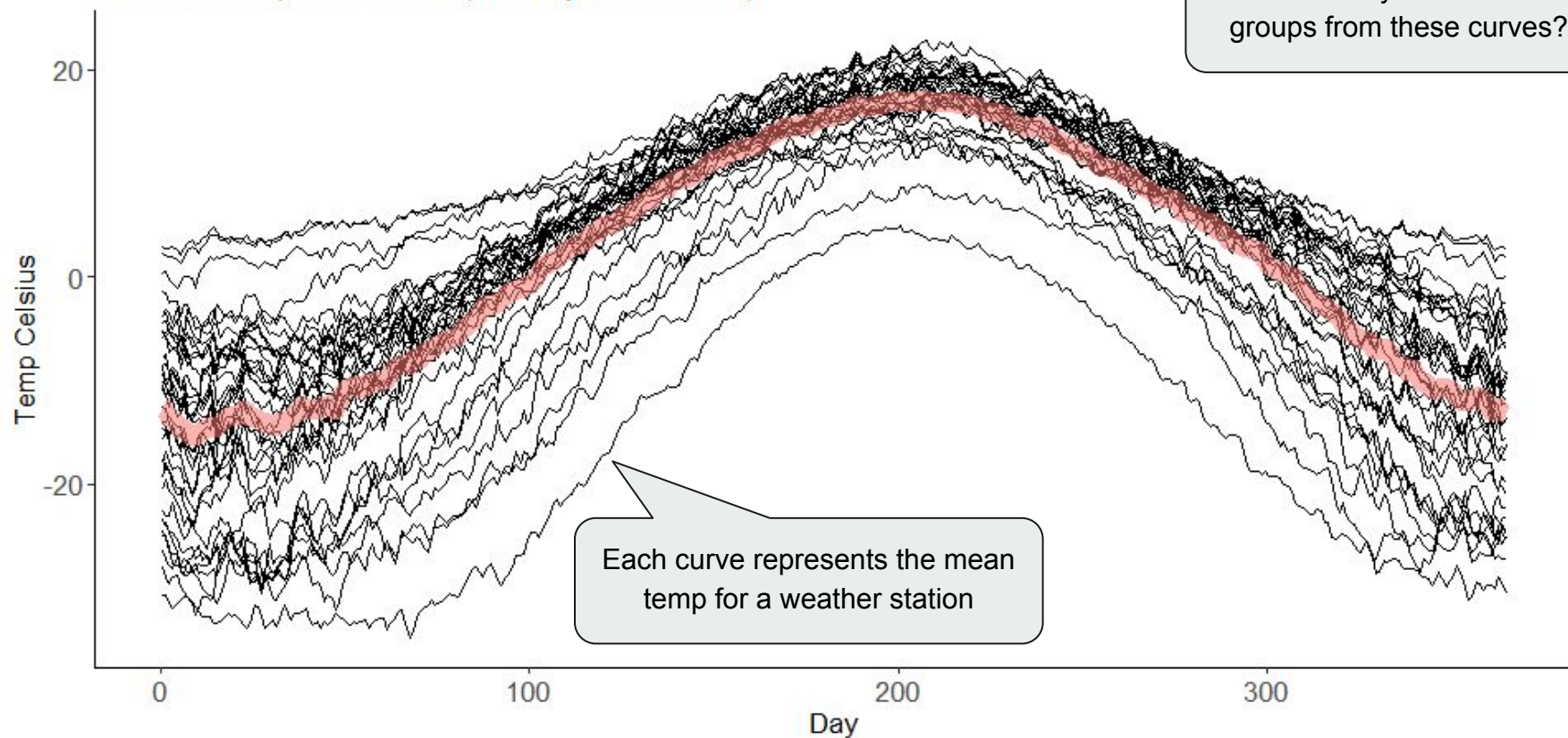
Pause



Have Questions?

Clustering | Segmentation

Canadian temperature data (Ramsay & Silverman)

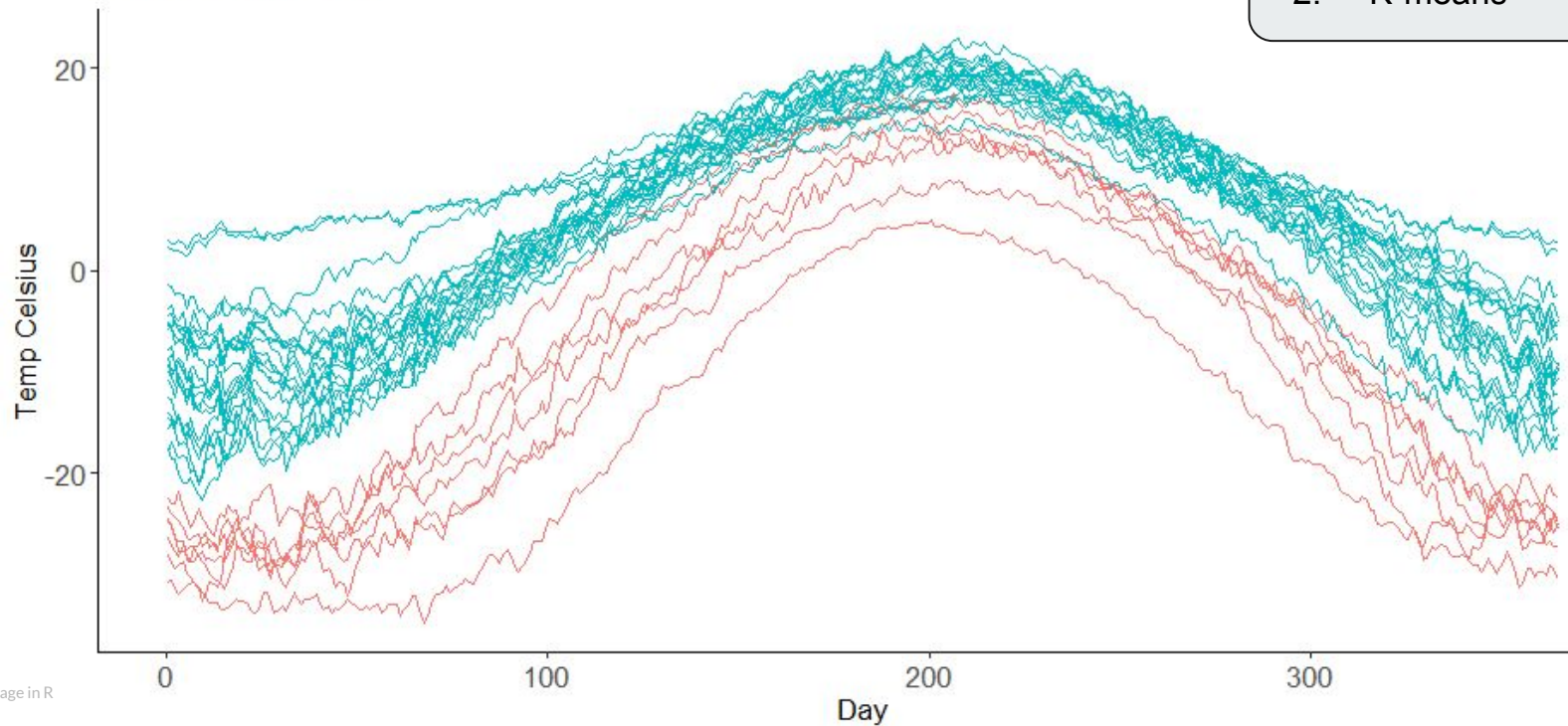


Traditional clustering methods

Canadian temperature data (Ramsay & Silverman)

Method = K-means, K = 2

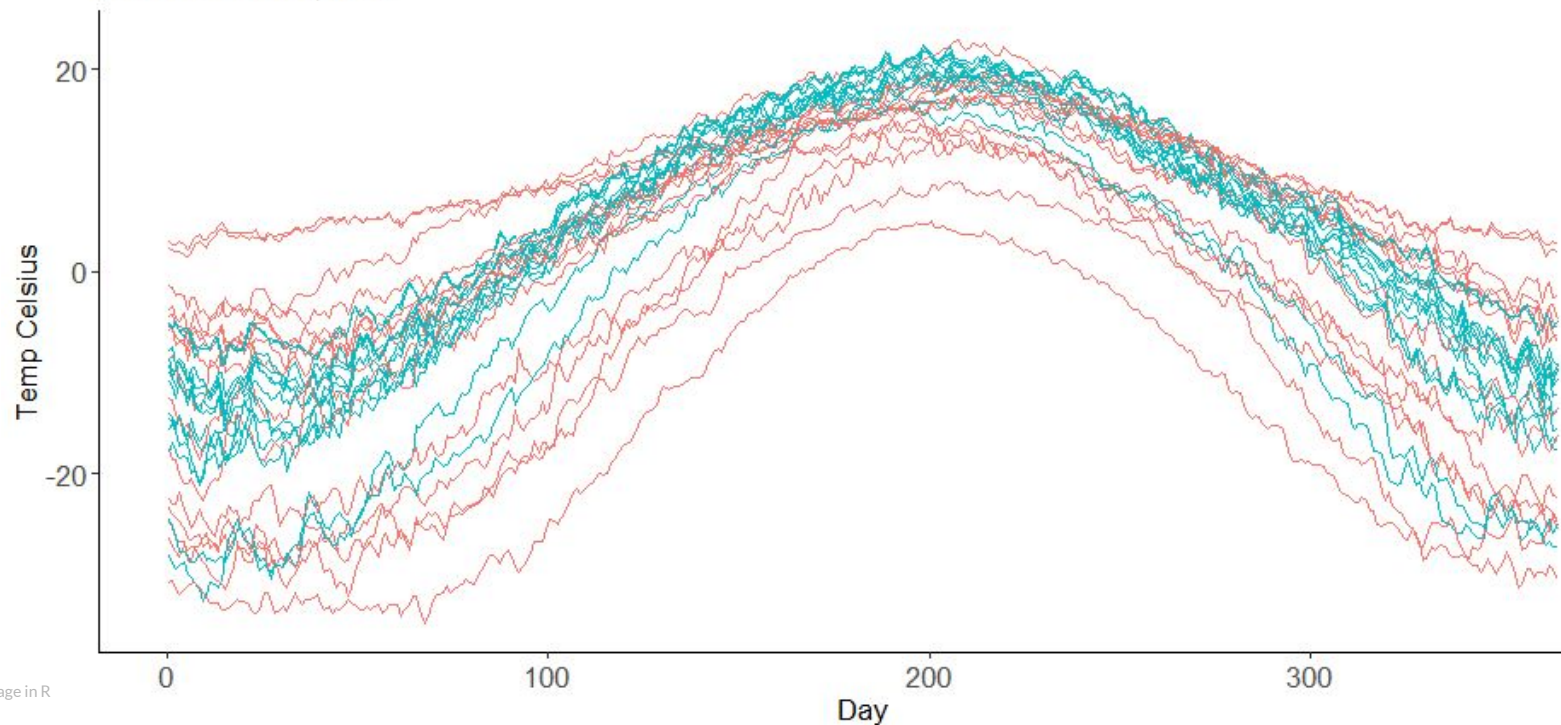
1. PCA (2 PCs)
2. K-means



Curve clustering using FDA

Canadian temperature data (Ramsay & Silverman)

Method = funHDDC, K = 2



Pause



Which clustering solution best
aligned with your thoughts?

Have Questions?





Resources for FDA

The End

- Active field of research
- Rich Resources in R:
 - Dedicated R CRAN Task View
 - Many great books
 - Functional Data Analysis by Ramsay and Silverman
 - Functional Data Analysis with R and Matlab by Ramsay, Hooker, and Graves
 - Blog posts such as Joseph Rickert's posts on R Views
 - Online course on FDA by Jiguo Cao (YouTube)
 - Online course: GAMs in R by Noam Ross¹
- There are Python resources too



1. <https://noamross.github.io/gams-in-r-course/>