Statistical and Sequential Learning for Time Series Forecasting

Introduction

Margaux Brégère

Course syllabus

- Framework
- Schedule
- Evaluation

Use case: forecasting french electricity consumption

- Challenges
- Data
- Descriptive analysis

Course syllabus

Time series forecasting framework

Let $Y = \left(Y_t \right)_{t \in \mathbb{N}^\star}$ be a discrete time processes where t refers to time and Y_t is a random variable

Assumption: at a time step $t = 1, 2, 3, \ldots$

- Observe the data with a delay d: $Y_1, Y_2, \ldots, Y_{t-d}$
- Observe potentially random explanatory variables $X_t \in \mathbb{R}^p$ - not restrictive since a categorial variable with m categories may be encoded in $\{1,\ldots,m\}$ or $\{0,1\}^m$

Aim

Providing forecasts $\hat{Y}_t, \ldots, \hat{Y}_{t+h}$ of the future realisations of Y at a horizon $h \in \mathbb{N}$

☞ Model $\hat{f}_t^h : \left(t, Y_1, Y_2, \ldots, Y_{t-d}, X_t\right) \mapsto \hat{Y}_{t+h}$ trained on $\left\{Y_s, X_s\right\}_{s=1,\ldots,t-d}$

Forecast evaluation:

On a testing dataset $\left\{Y_s, X_s\right\}_{s=T_1,\ldots,T_n}$ and a loss function ℓ, we aim to minimise

$$\frac{1}{n} \sum_{s=T_1}^{T_n} \ell\left(Y_{s+h}, \hat{Y}_{s+h}\right)$$

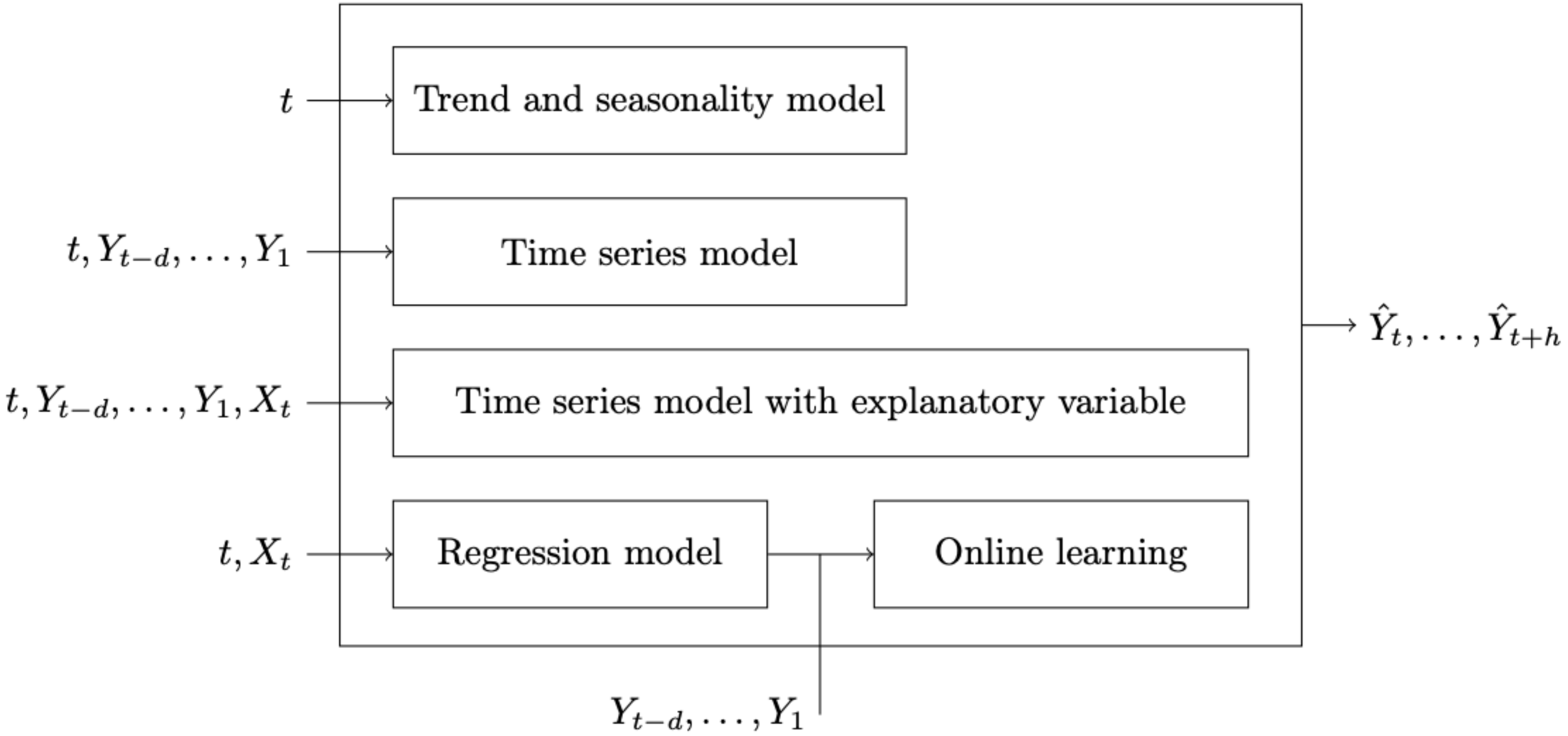
Time series forecasting framework

Supervised (target Y_t)

Sequential $\left(\hat{f}^h_{t+1} \neq \hat{f}^h_t \right)$ learning

for Regression ($Y_t \in \mathbb{R}$)

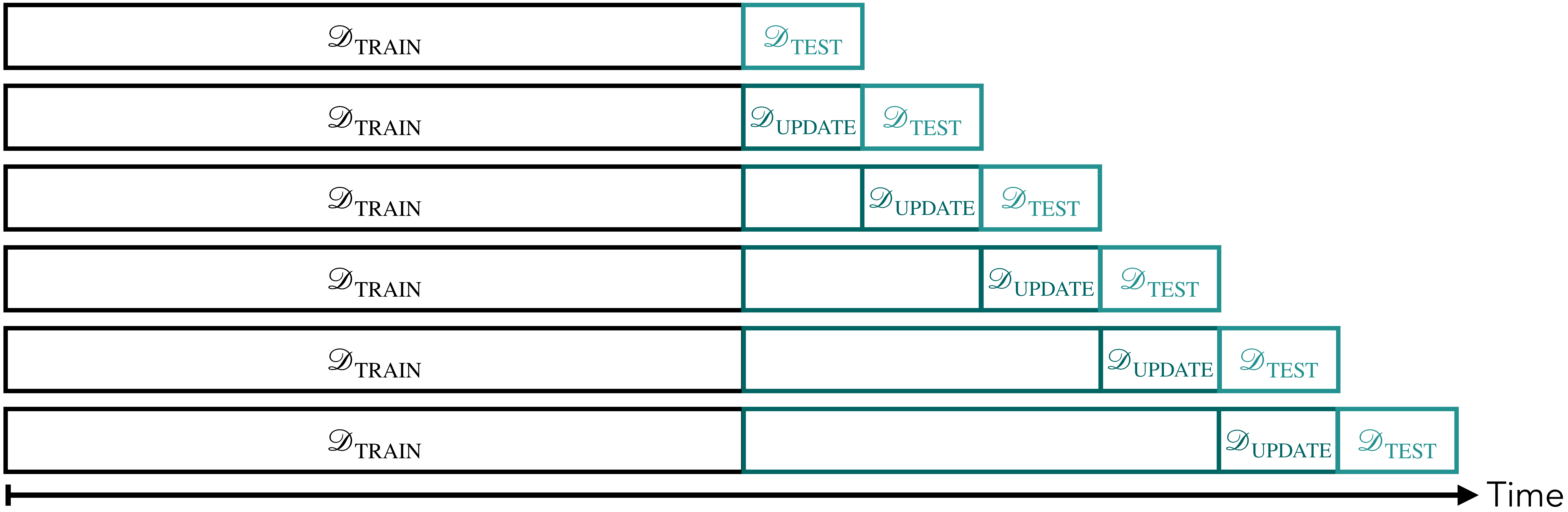
Time series forecasting



Time series forecasting

- Classical learning

- Online / sequential learning



Objectives

- Understand the challenges involved in forecasting time series:

    Mean forecast, probabilistic forecast, quantiles, extremes, simulation

    Online learning (adaptive models, transfer learning, expert aggregation)

- Present various statistical and learning methods (linear, random forests, bagging, boosting, neural networks, etc.)

- Implement these methods in R and/or Python as the course progresses

Schedule

Oct. 23, 2023 09h00 - 12h15 + 17h00 - 18h30 • Introduction, Times series analysis

Oct. 25, 2023 13h30 - 16h45 • Penalised Regression, Online approaches

Nov. 06, 2023 09h00 - 12h15 + 17h00 - 18h30 • Generalised Additives Models, Online approaches

Nov. 08, 2023 13h30 - 16h45 • Random Forest and Boosting, Online approaches

Nov. 15, 2023 13h30 - 16h45 • Project session

Nov. 20, 2023 09h00 - 12h15 • Recurrent Neural Networks and variations

Nov. 22, 2023 13h30 - 16h45 • Online expert aggregations

Nov. 29, 2023 13h30 - 16h45 • Interpretability

Dec. 06, 2023 13h30 - 16h45 • Project session

Dec. 13, 2023 13h30 - 16h45 • Opening: probabilistic forecasting and generative models

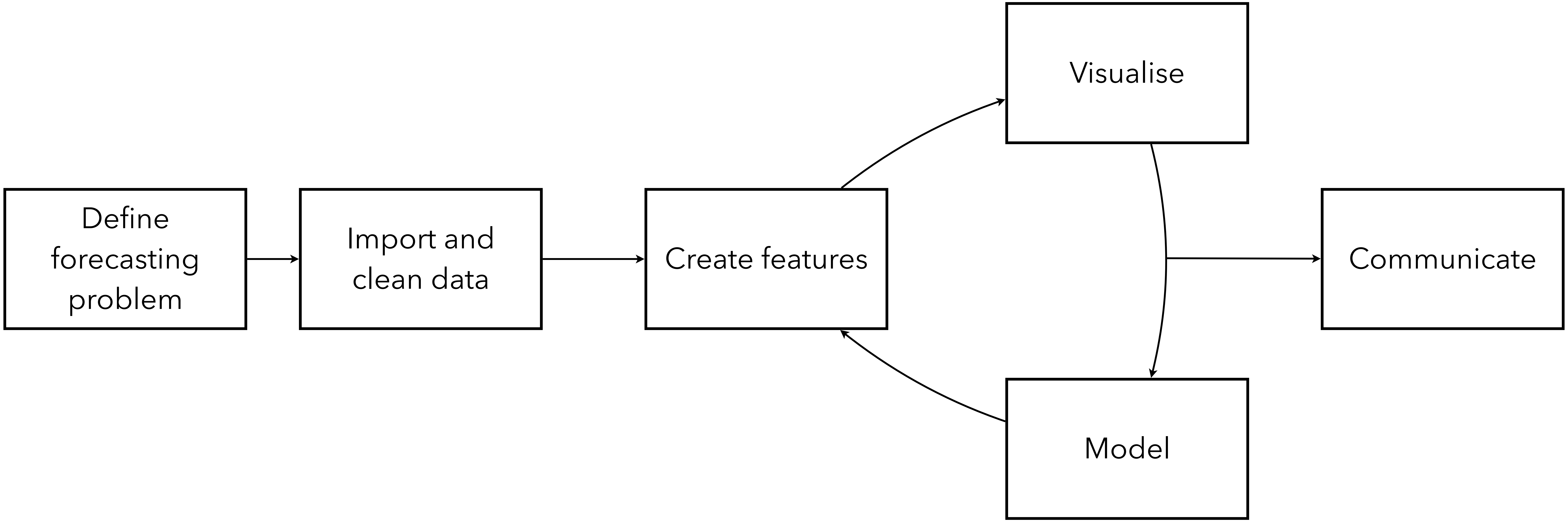
Evaluation - Project timeline

- Constitute pairs or trios
- Choose a dataset (data.gouv.fr, kaggle.com/datasets, data.oecd.org, etc.)

  A single requirement: the variable of interest must be a time series!
- Define our forecasting problem (motivation, delay, horizon)

  ☑ You need my approval to continue
- Define benchmark forecasts (persistent model, existing predictor)
- Define a testing methodology
- Work: impute missing data, design features, model, visualise, tune to beat the benchmark
- Do not forget sequential aspect (online learning)
- Provide reproductible results, analyse

Project



« All models are wrong, but some are useful »

George Box

Objectives

Handle real data

Pose a forecasting problem

Understand statistical and online learning methods and their implementation (R or python)

Work in pairs/trios and use collaborative tools (Git)

Build and validate a predictive model

Present results orally and in writing

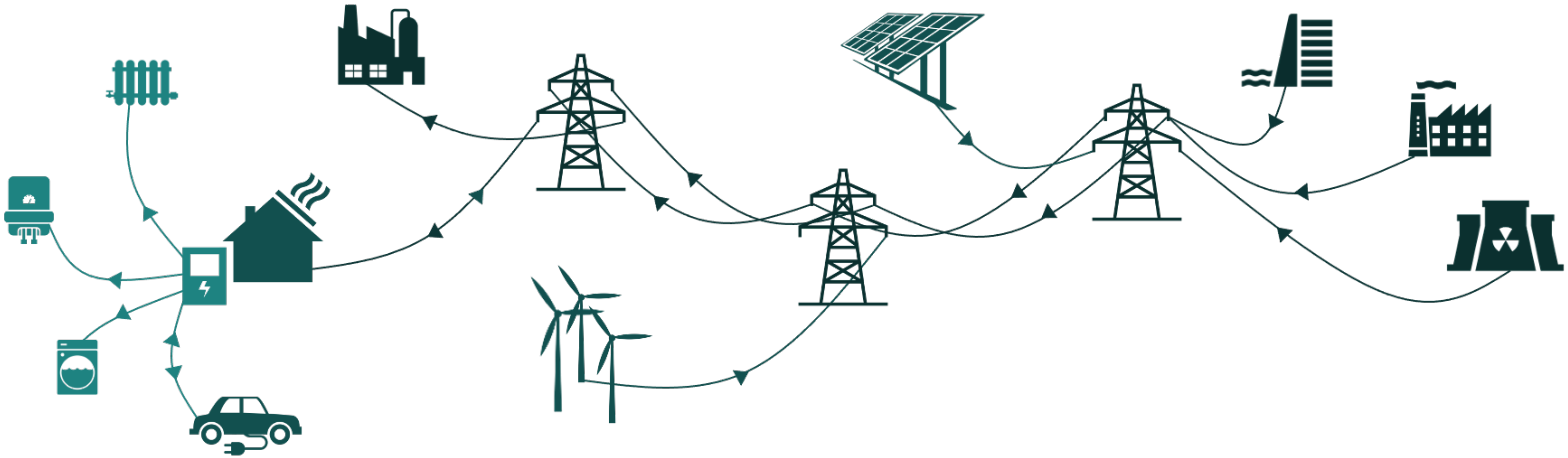
Provide reproducible code

Yardstick of the evaluation

1 point	• Literature review
2 points	• Data processing (imputation of missing data, feature engineering)
1 point	• Descriptive analysis
5 points	• Modelling and validation of final model (test methodology)
2 points	• Analysis of results (bias, variance, etc.)
2 points	• Code (comments, reproducibility, effectiveness)
3 points	• Report
4 points	• Oral presentation (fluency and clarity, motivation, pedagogy, answers to questions)

Use case: forecasting french electricity consumption

Challenges



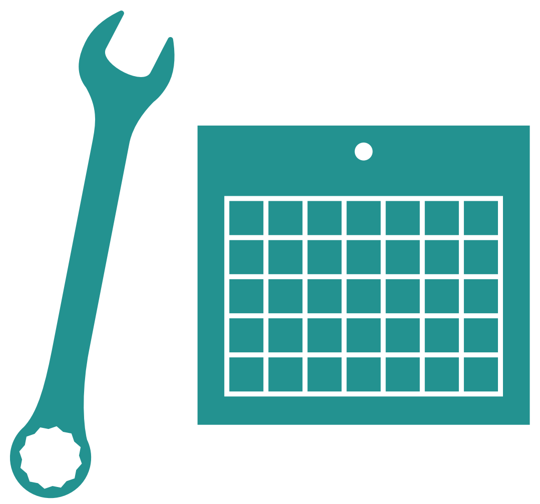
As electricity is hard to store, balance between production and demand must be strictly maintained

Forecast demand and adapt production accordingly

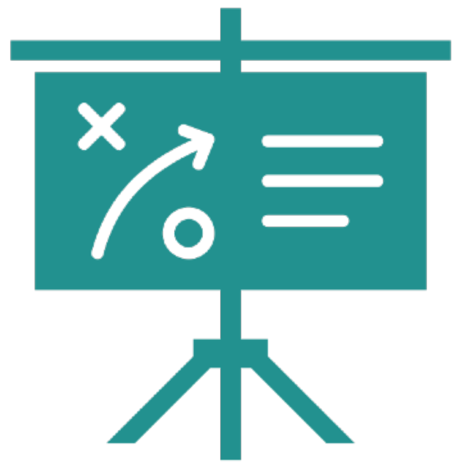
From short to long-term

- Short term (from a few hours to two weeks)
  Scheduling and optimising the use of power plants
  Avoiding black outs
  Reducing rebalance coasts

- Mid term (from two weeks to five years)
  Planning power plant maintenance episodes.



- Long term (from five to fifty years)
  Providing prospects for the evolution of the customer portfolio
  Adapting commercial offers accordingly
  Defining an investment strategy



From dis-aggregated to aggregated level

- Dis-aggregated level
  Modelling new electrical uses (auto-consumption, electrical vehicles)
  Designing demande response solutions
  ⚠ Smart meters data is highly sensitive and erratic ☞ simulation models

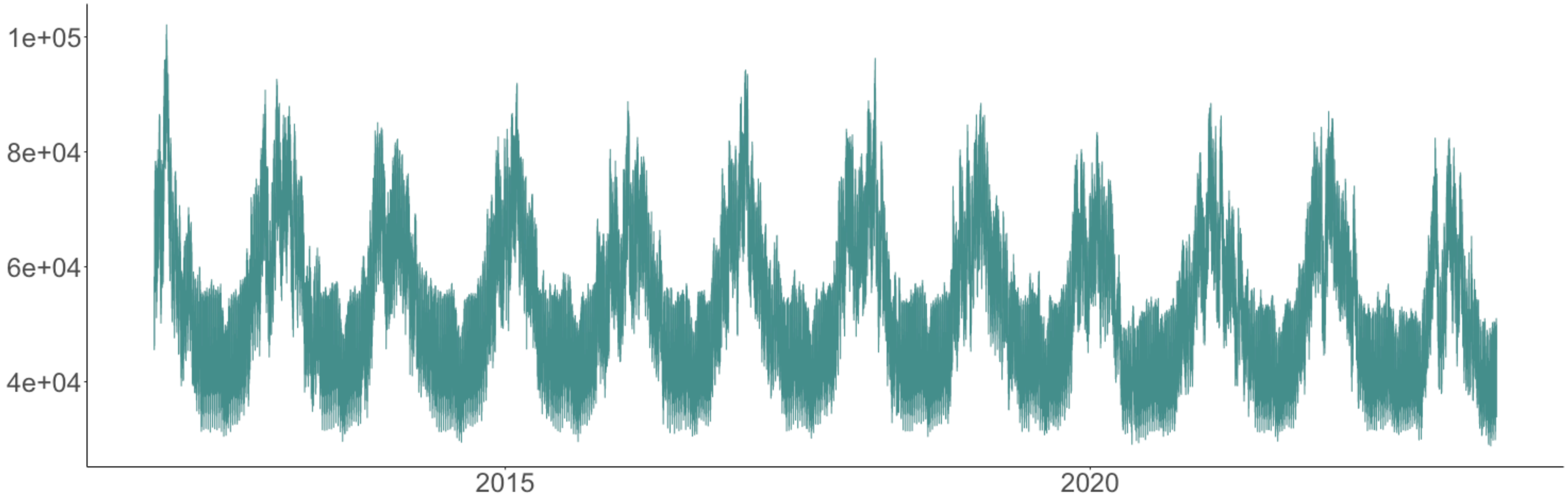
- Neighbourhood / city level
  Managing networks locally (Smart Grids)
  Dispatching electricity at junctions between transport (high-voltage lines) and distribution (medium- and low-voltage lines) networks

- National level
  Managing the overall balance
  Planning cross-border exchanges

Growing area of research: forecast reconciliation

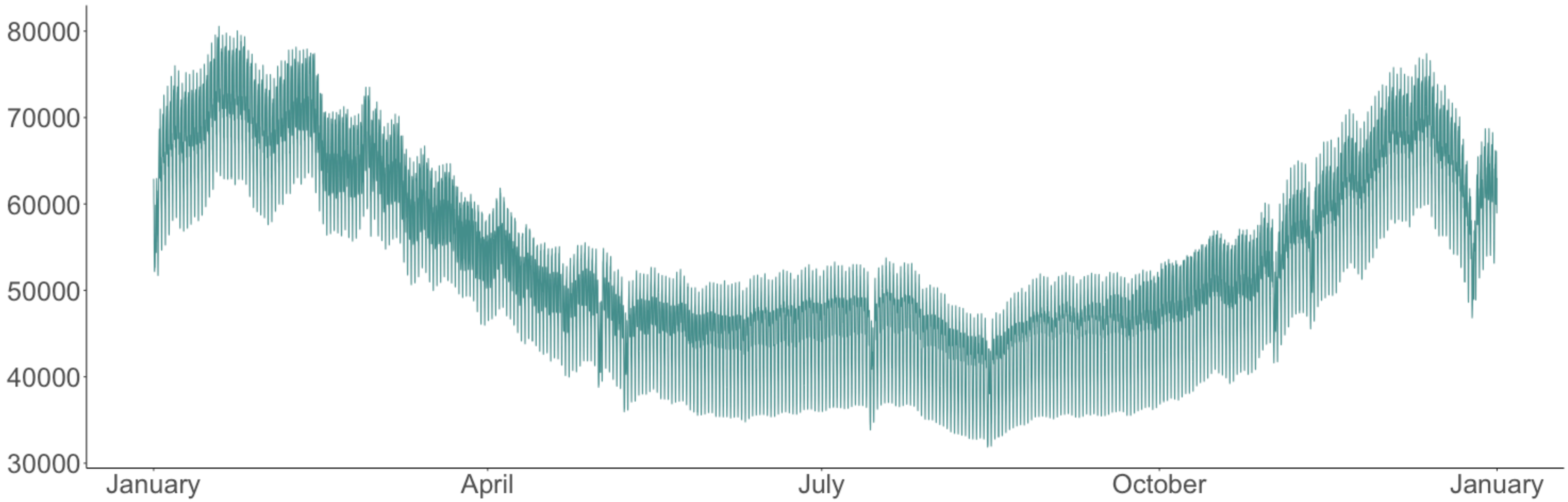
Data

https://www.rte-france.com/eco2mix

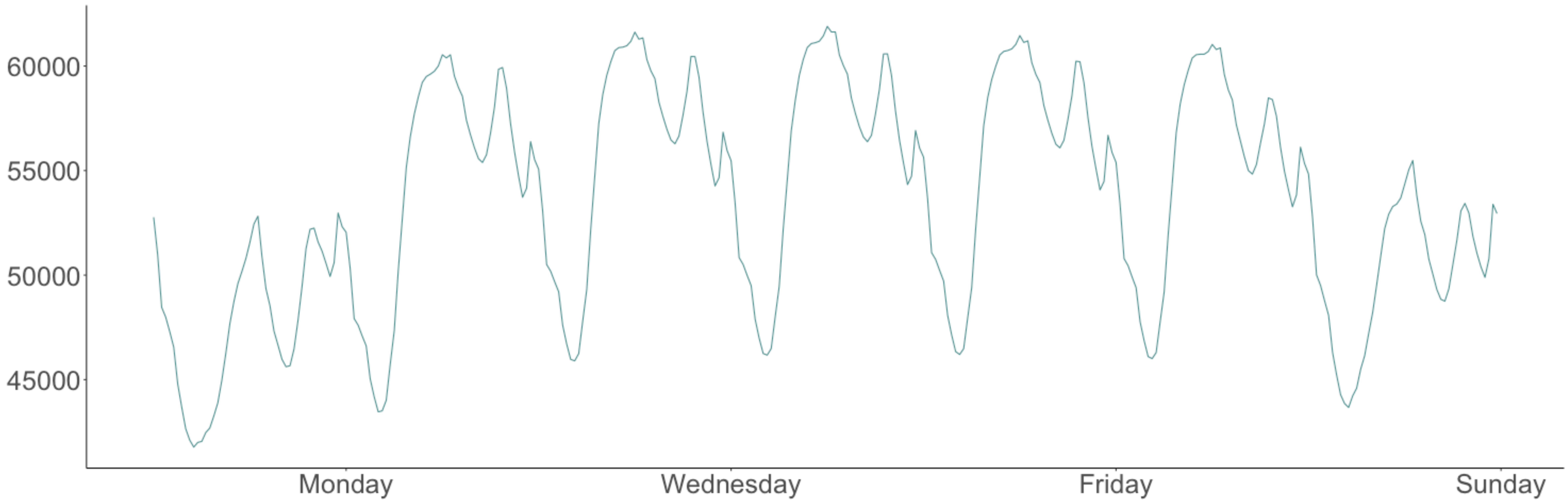


French electricity consumption (MW) in half-hourly time steps
from January 1, 2012 to June 22, 2023

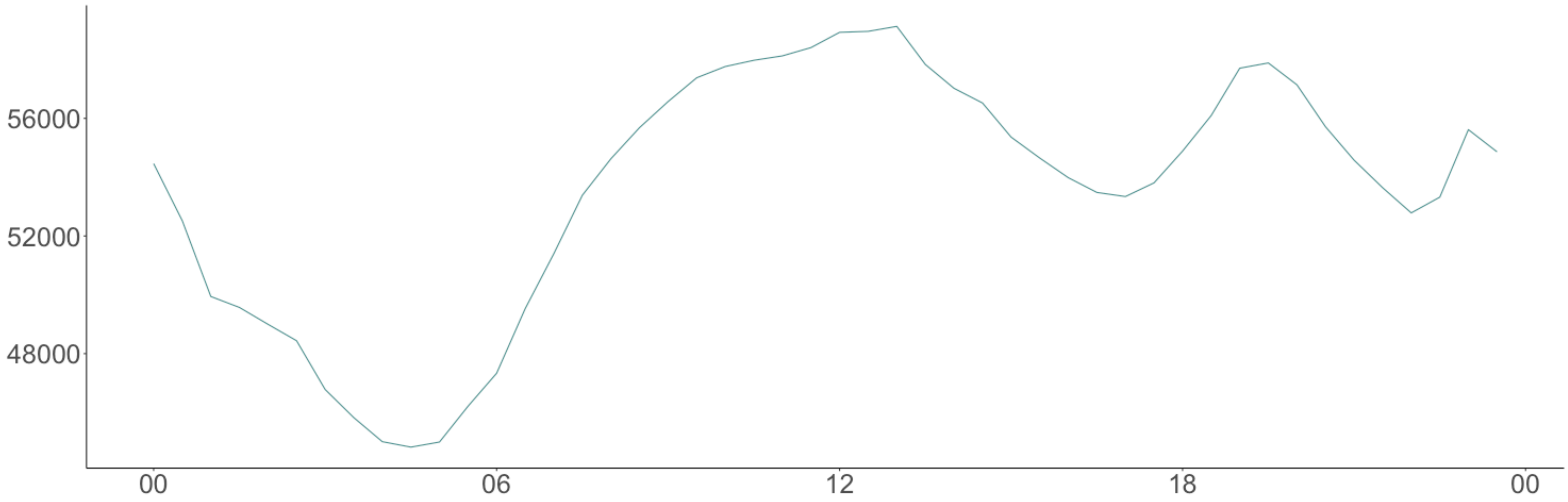
Descriptive analysis - average annual load



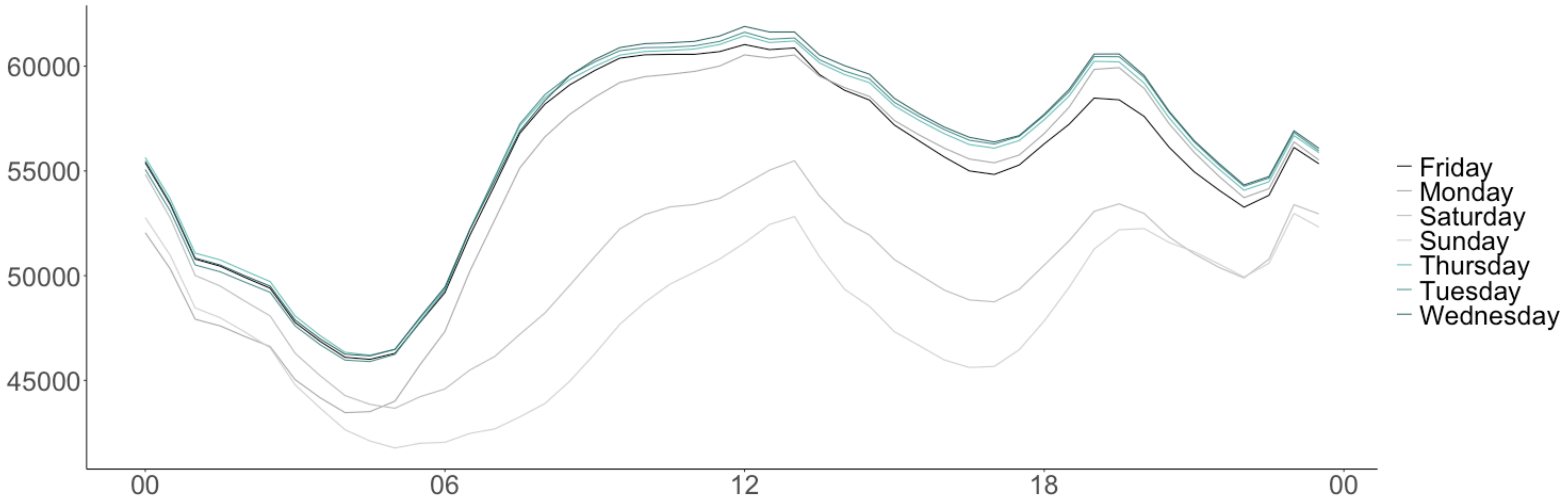
Descriptive analysis - average weekly load



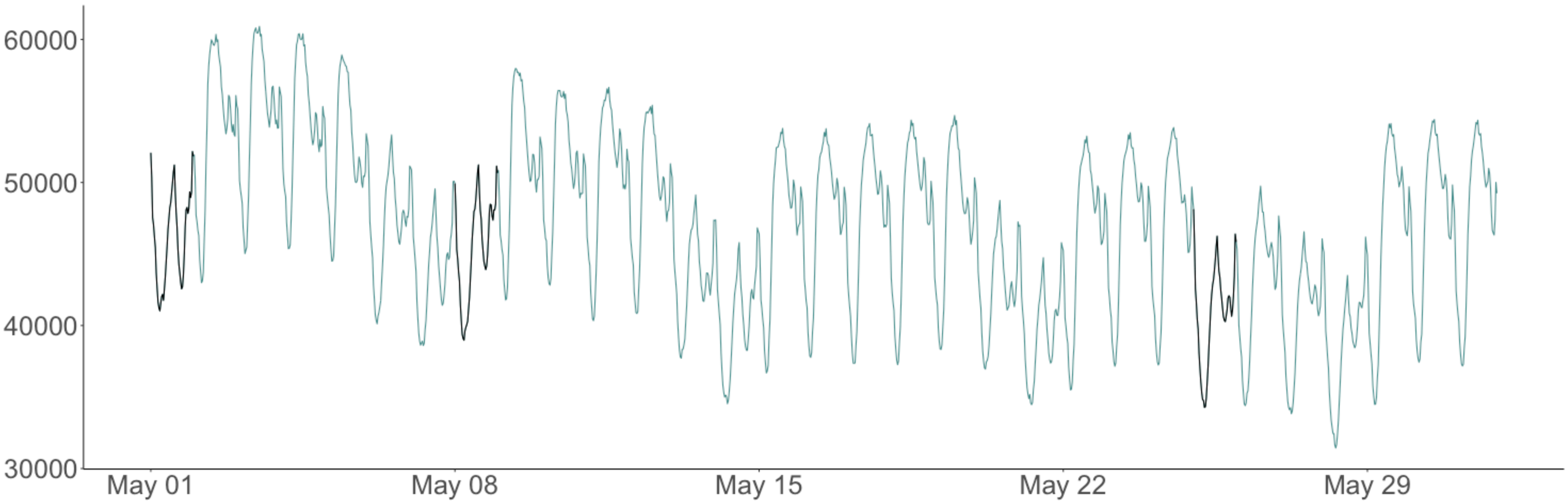
Descriptive analysis - average daily load



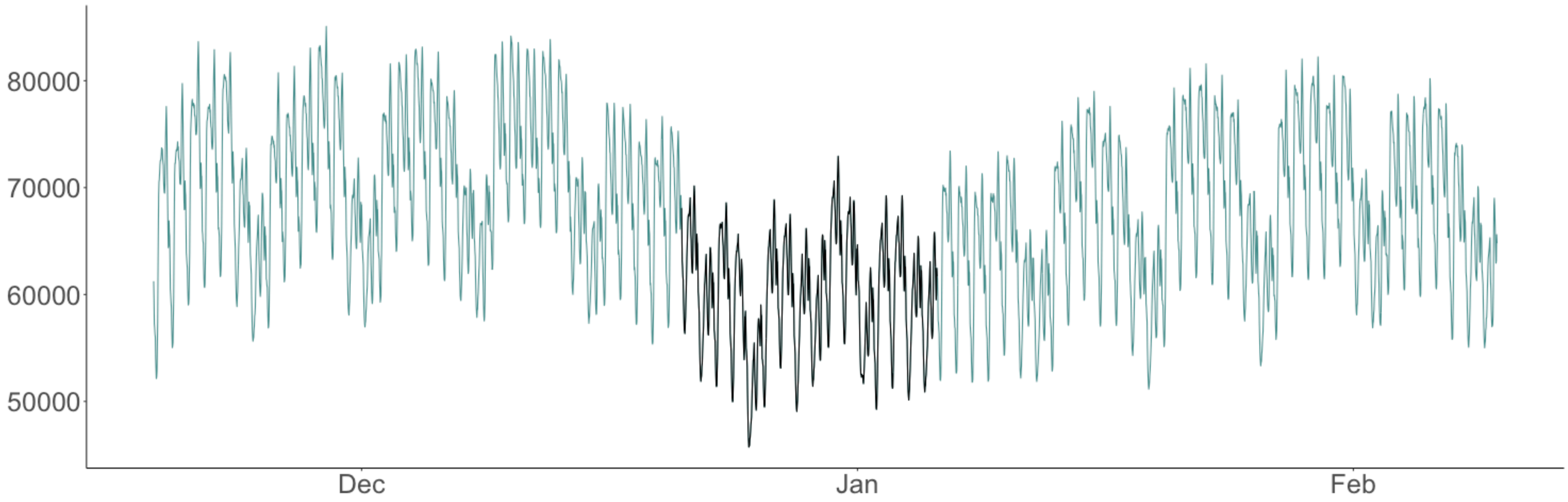
Descriptive analysis - type of day



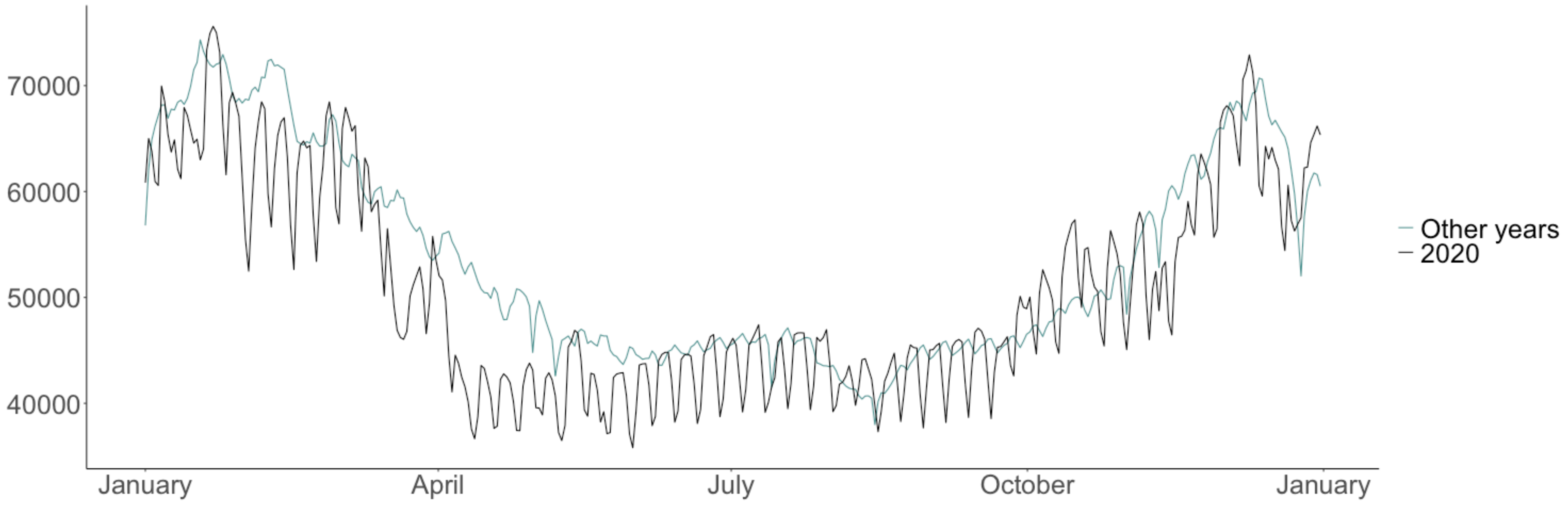
Descriptive analysis - day off



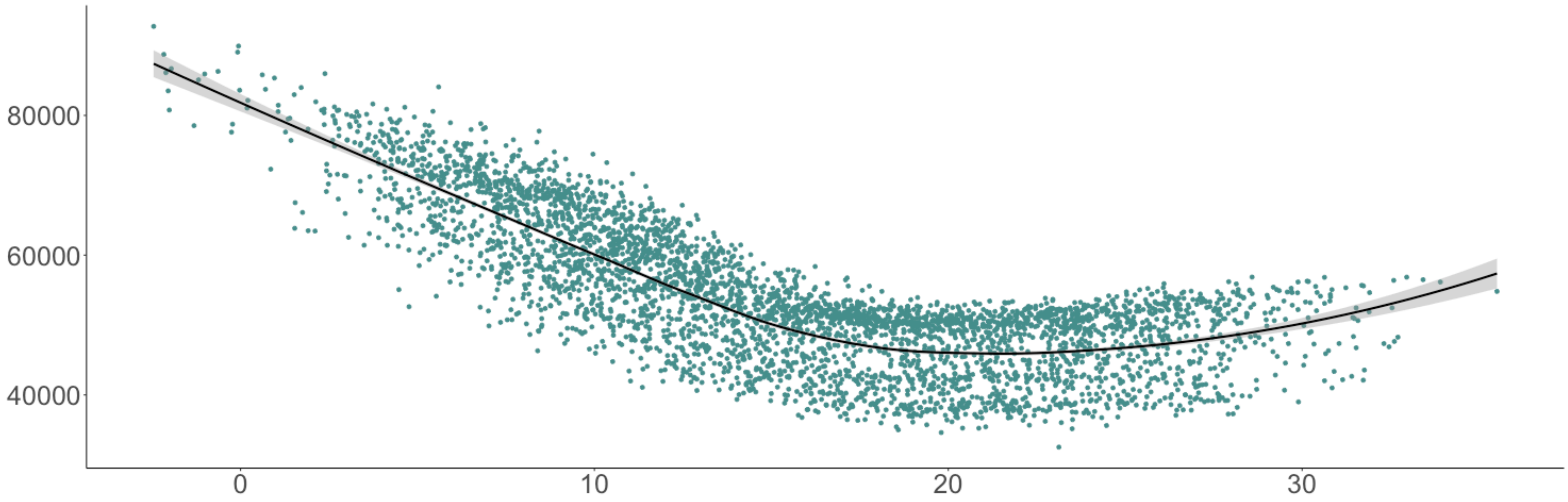
Descriptive analysis - holidays



Descriptive analysis - rare event



Descriptive analysis - temperature impact at 4p.m.



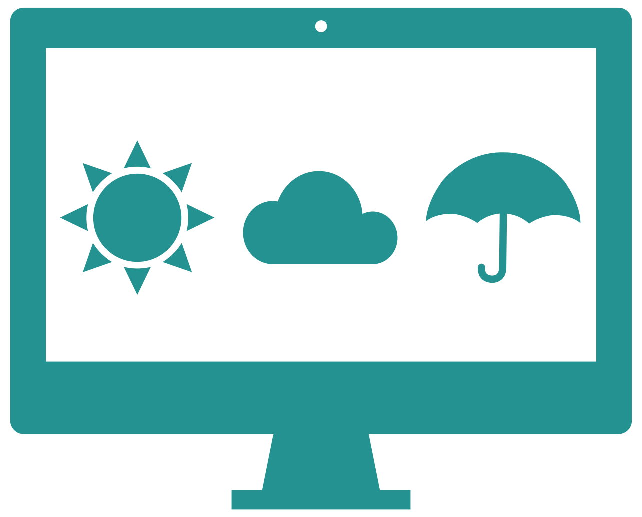
Additional data and assumption

https://donneespubliques.meteofrance.fr/

39 weather stations in mainland France (excluding Corsica)

Observations at three-hour time step of:

- Temperature,
- Nebulosity,
- Wind Speed,
- Humidity,
- Precipitation etc.



Météo France has forecast the weather perfectly, so we will use the observations as if they were forecasts

That's all folks!

TP → https://drive.google.com/drive/folders/1OH6oH0dolXN9QxOSKuEfLbF44HkN5QfB?
usp=share_link