ENERGY DATA SCIENCE

Data processing: Feature engineering

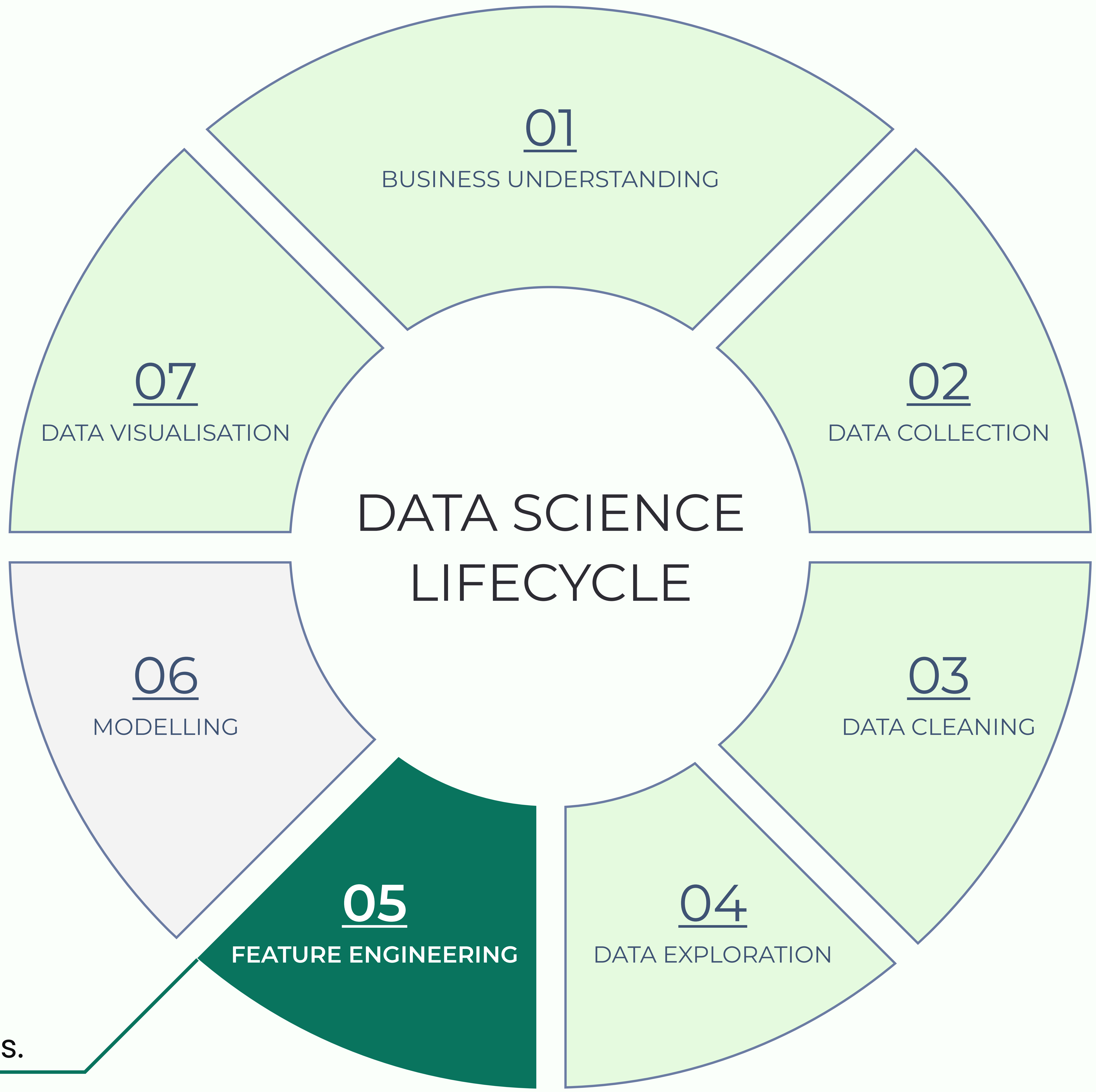
Prof. Juri Belikov

Department of Software Science
Tallinn University of Technology
juri.belikov@taltech.ee

PREVIOUSLY IN ITS8080 …

Key takeaways:

- Data types
- Basic data cleaning steps
- Missing data mechanisms
- Handling missing data:
  - Deletion
  - Uni- and multivariate imputation

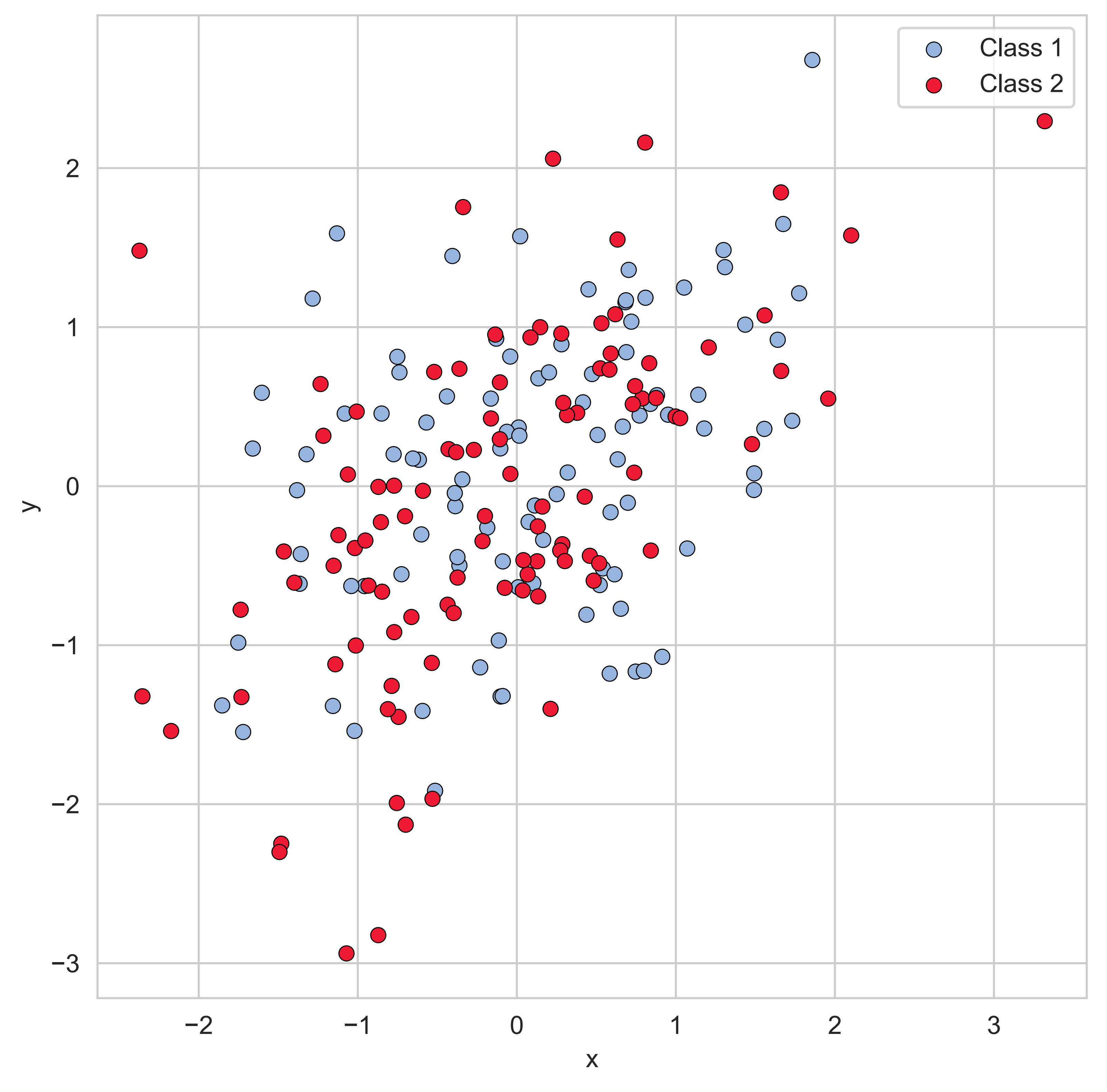


Select important features and
construct more meaningful ones.

Illustrative Examples

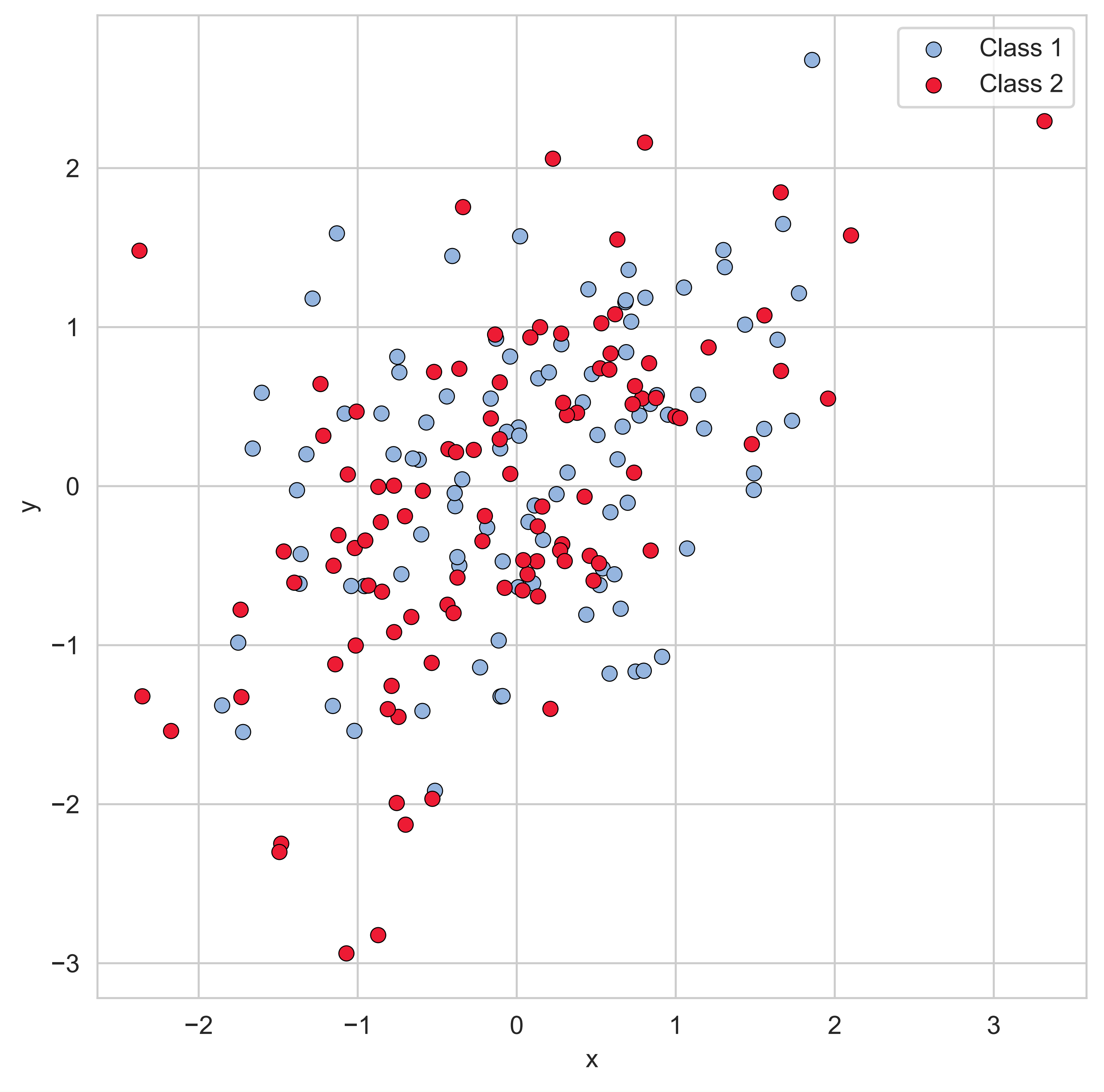
SYNTHETIC EXAMPLE

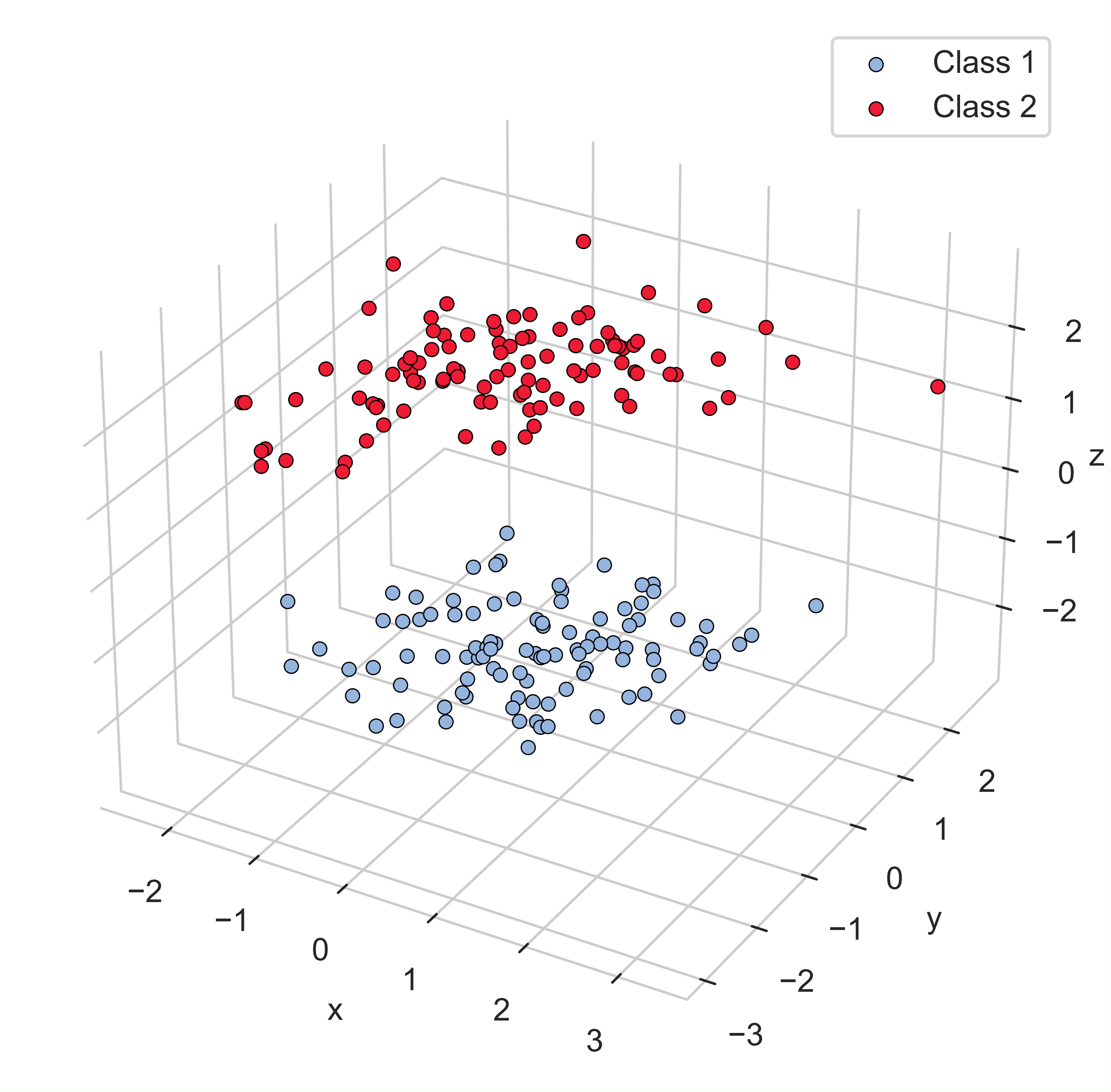
Any thoughts?



SYNTHETIC EXAMPLE

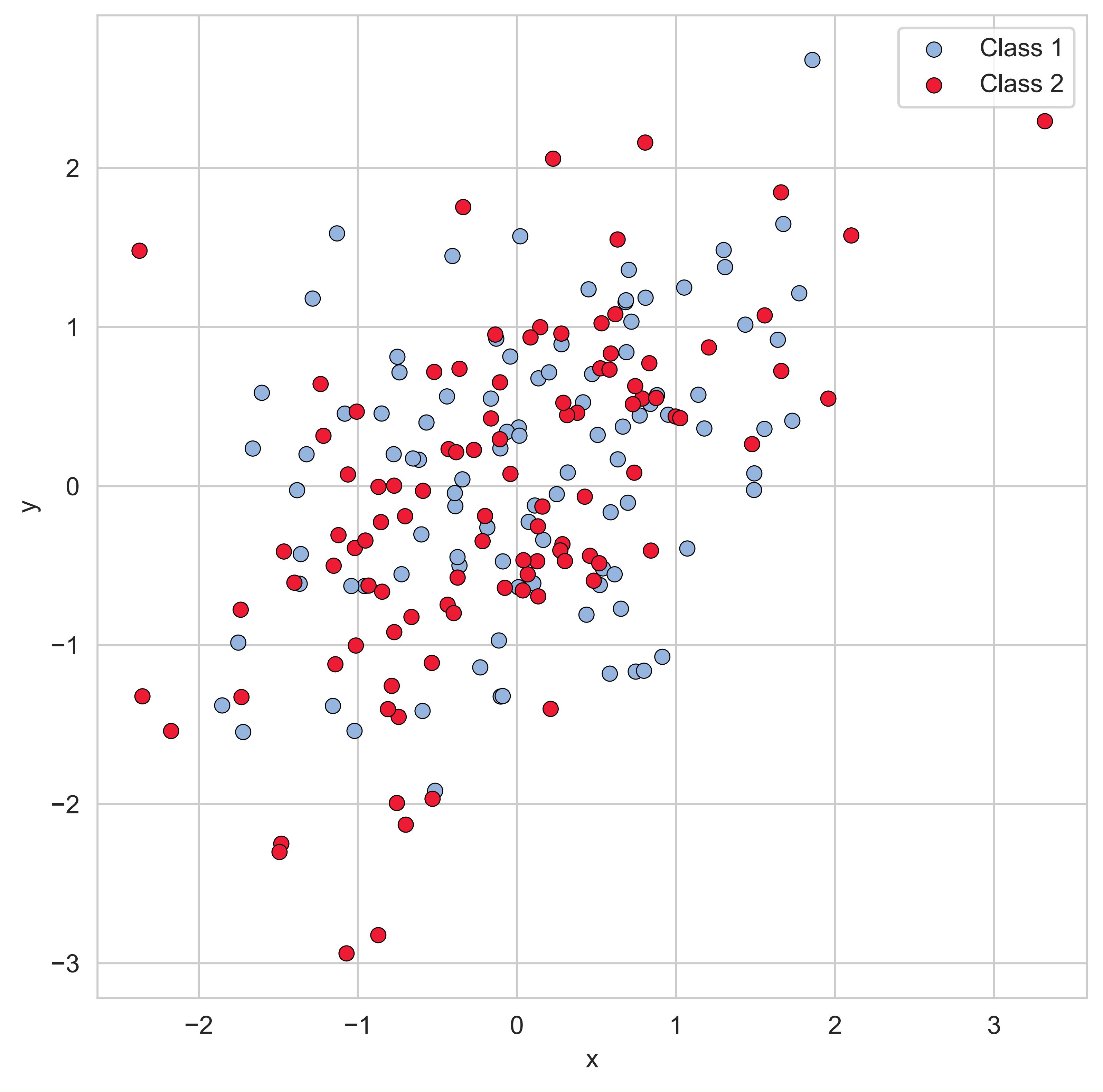
Any thoughts?



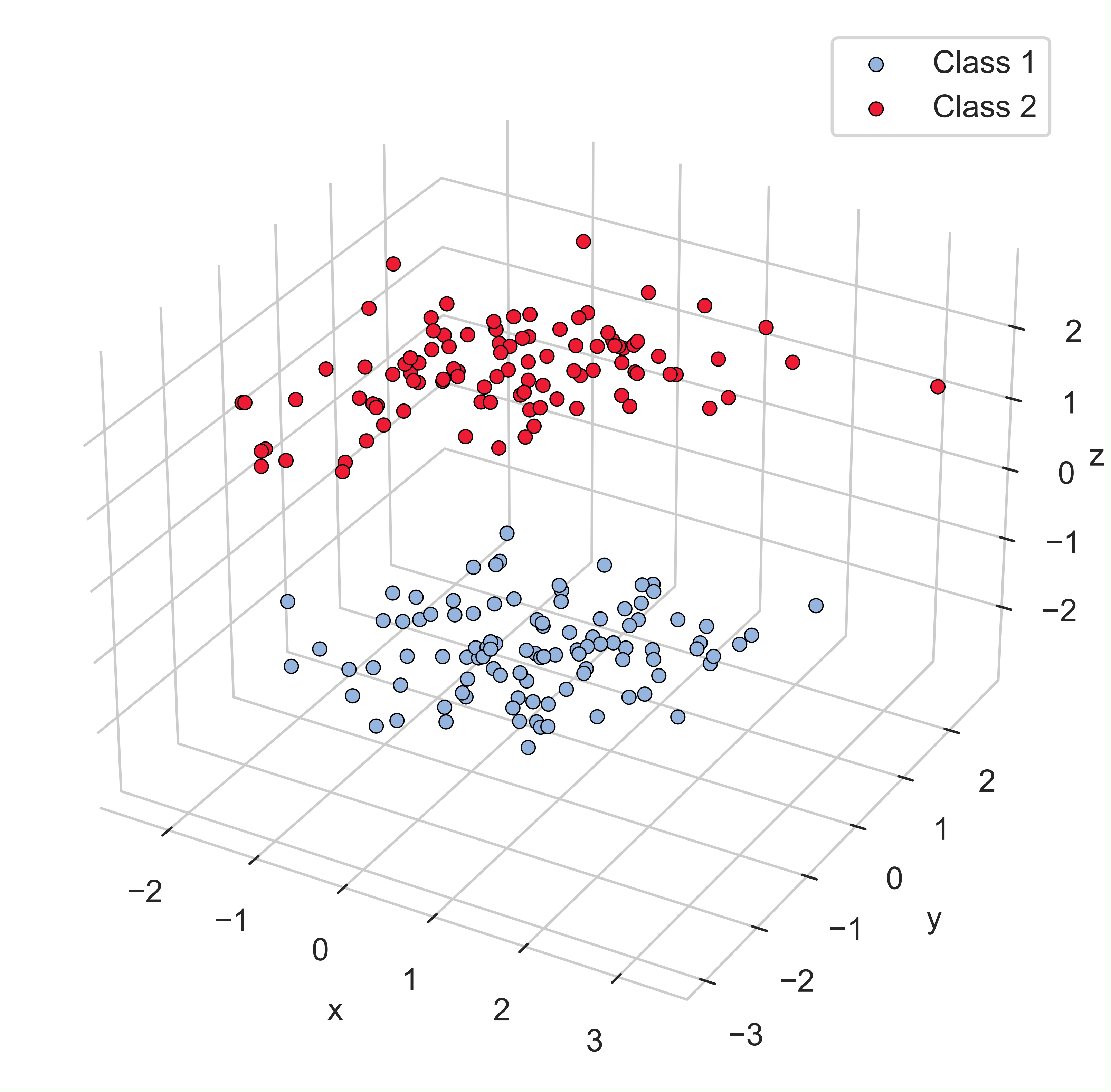


SYNTHETIC EXAMPLE

Any thoughts?



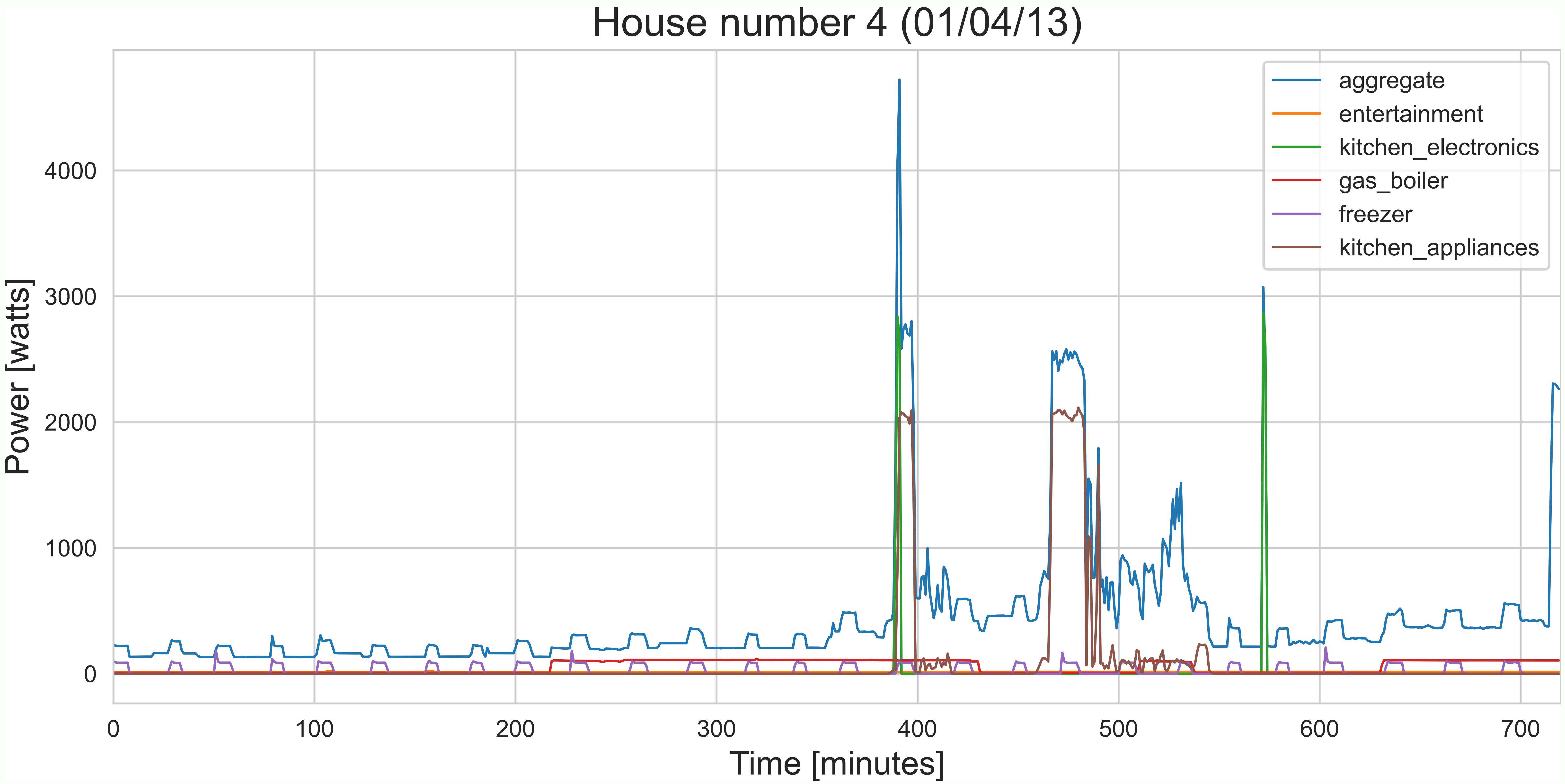
$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$



$$z \sim \mathcal{N}(\pm 2, 0.09)$$

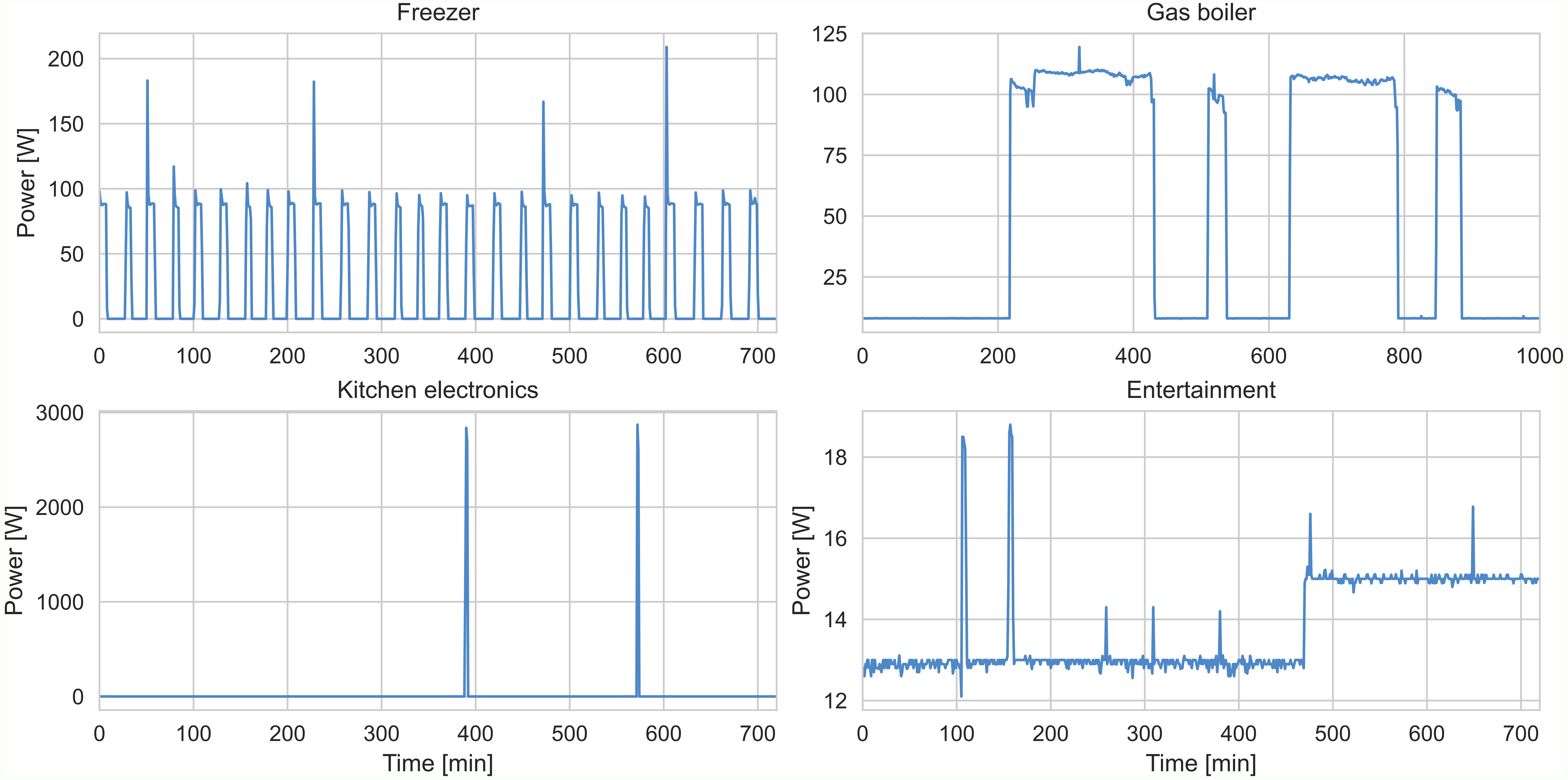
REAL DATASET

Looks messy when plotted together.



SELECTED APPLIANCES

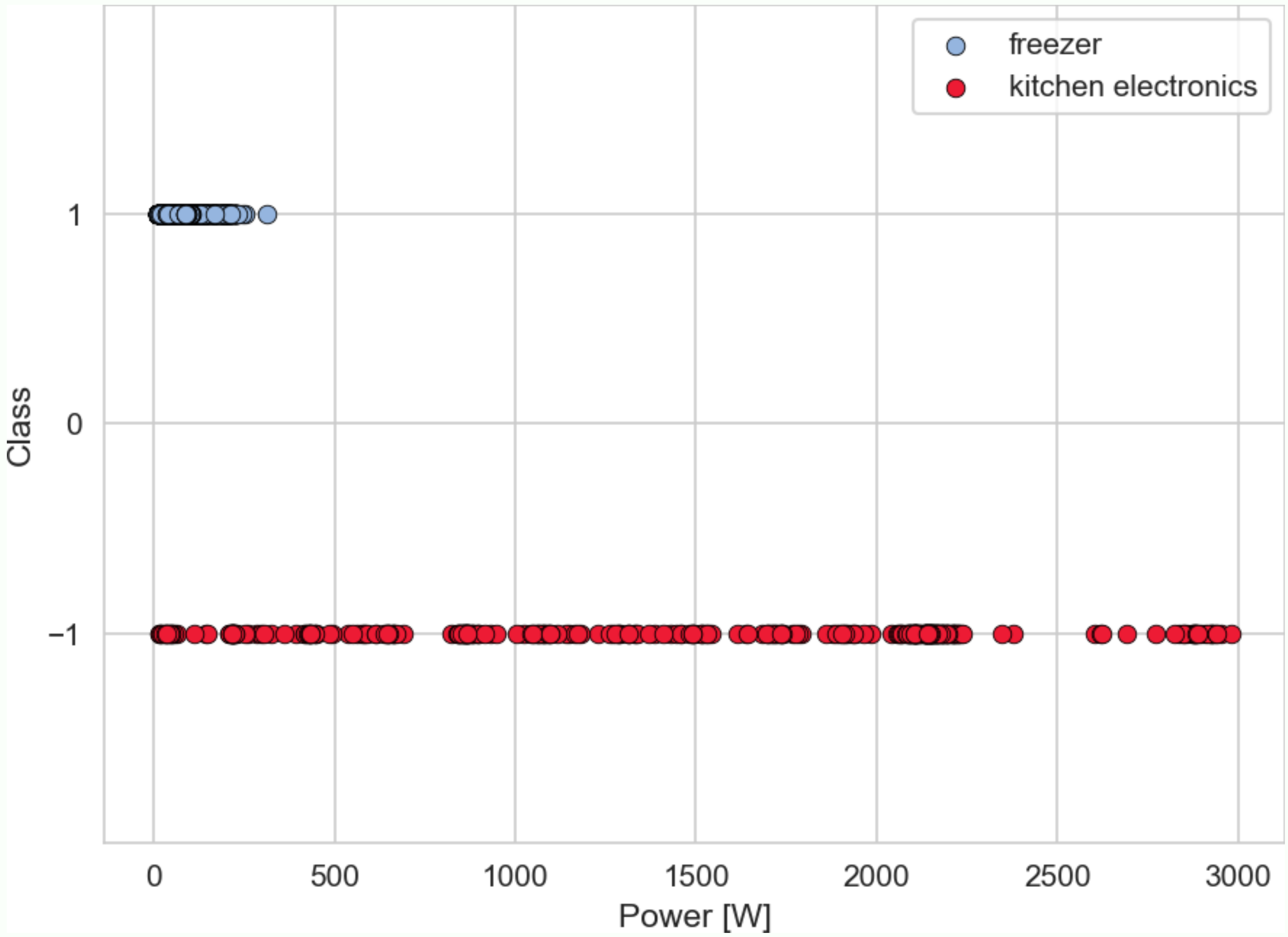
Looks better when plotted separately.



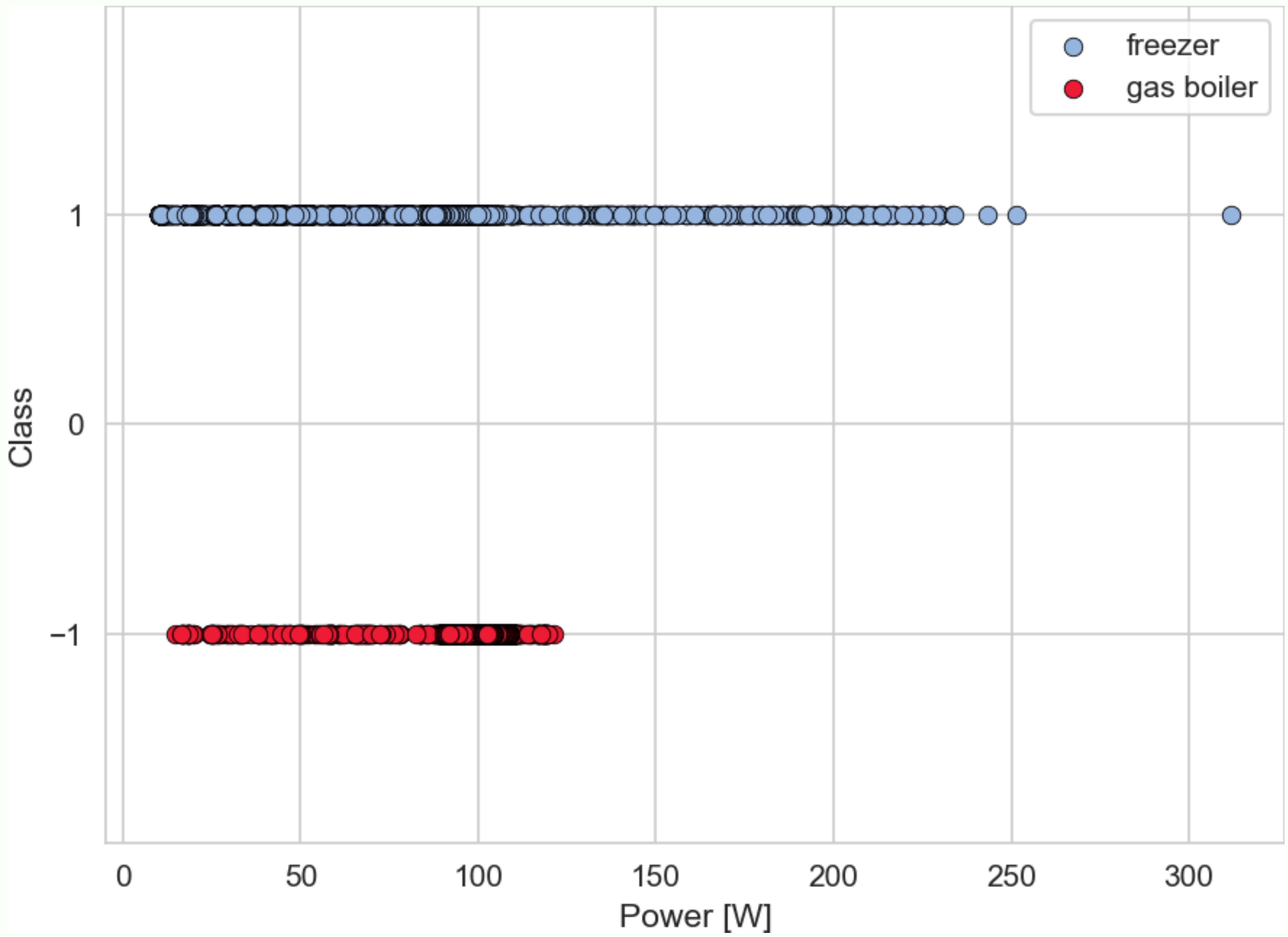
Our goal is to understand which appliance is working at any given moment.

1D CASE: GEOMETRIC VIEW

easy

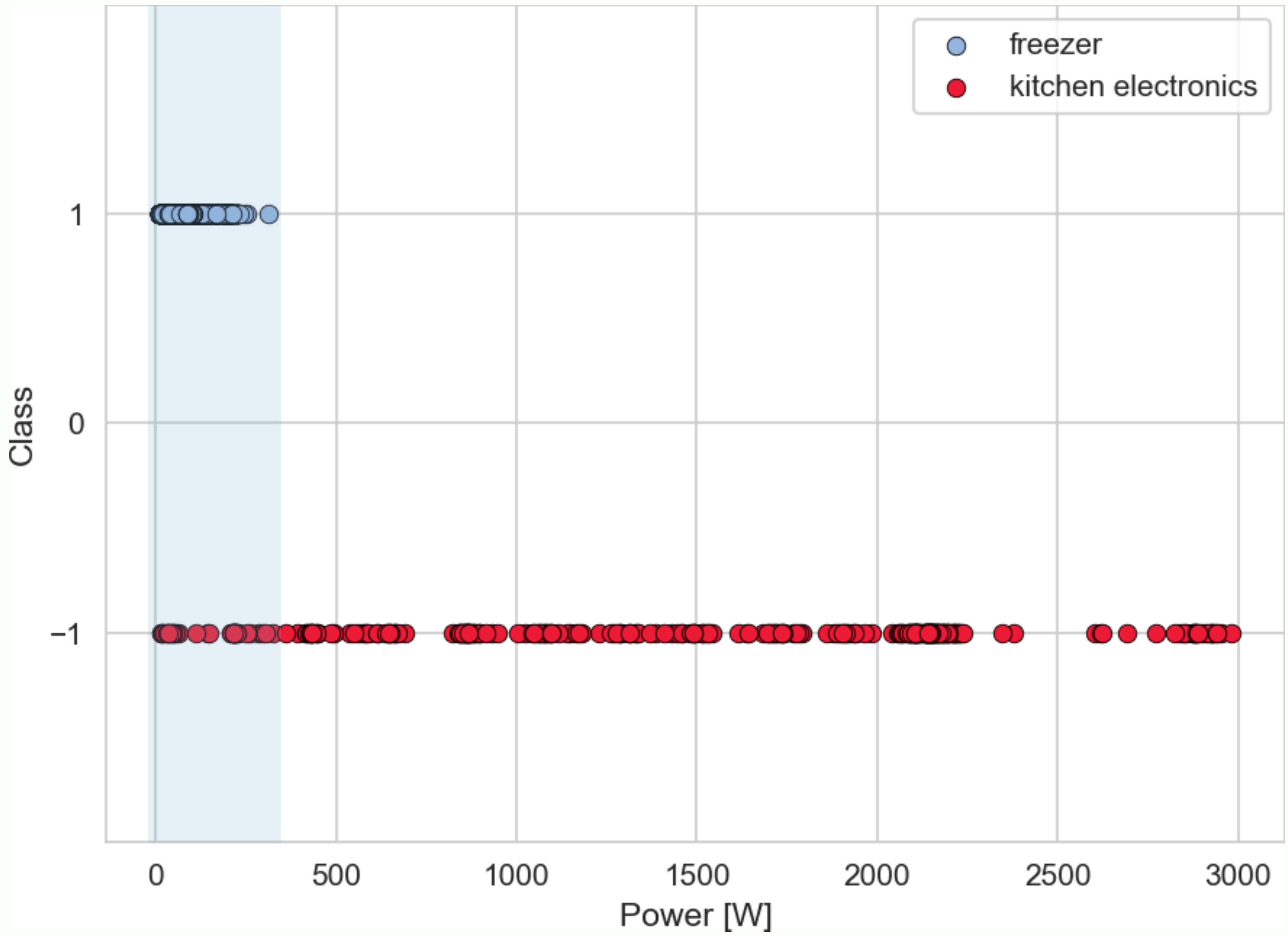


hard

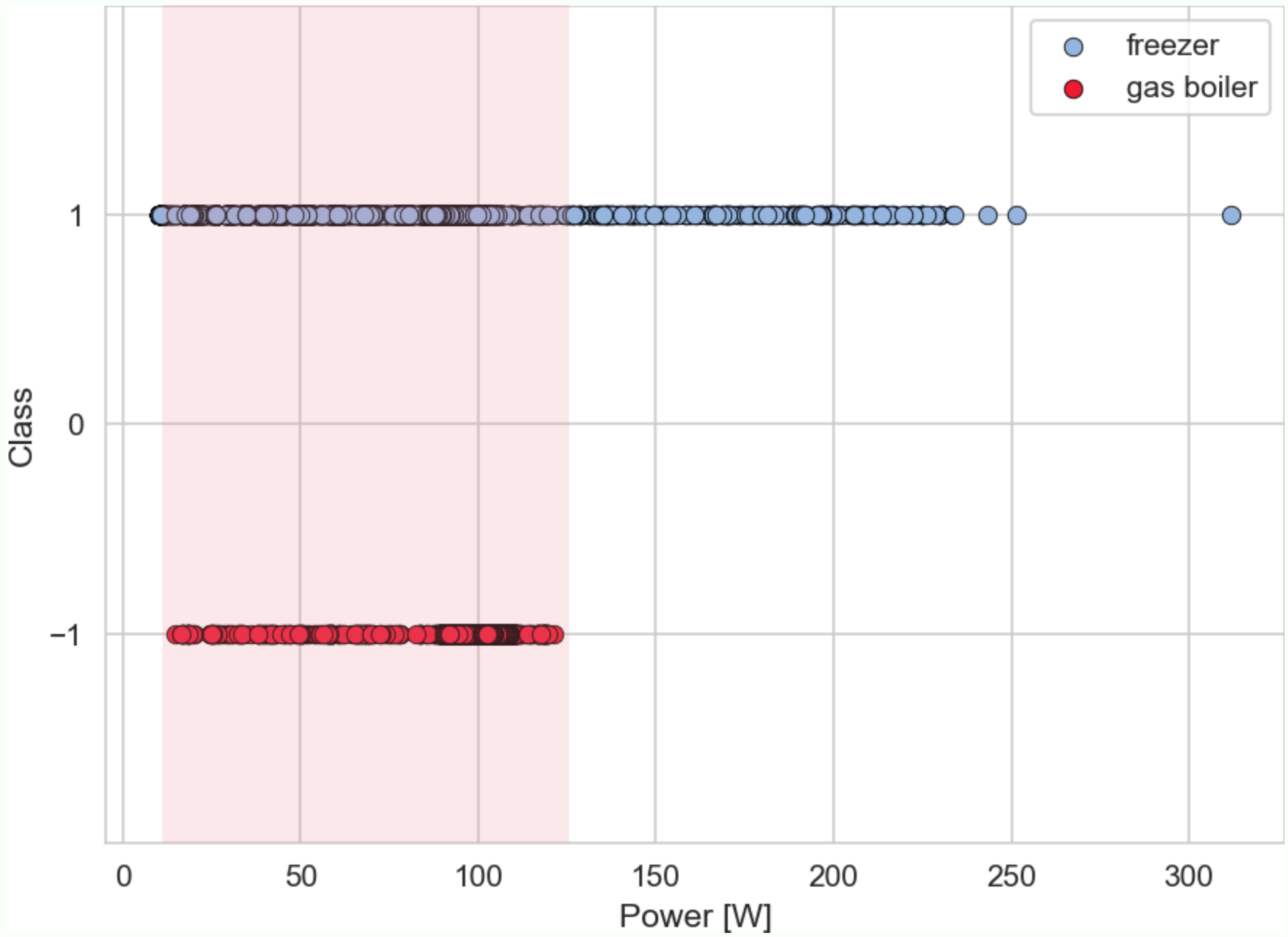


1D CASE: GEOMETRIC VIEW

easy

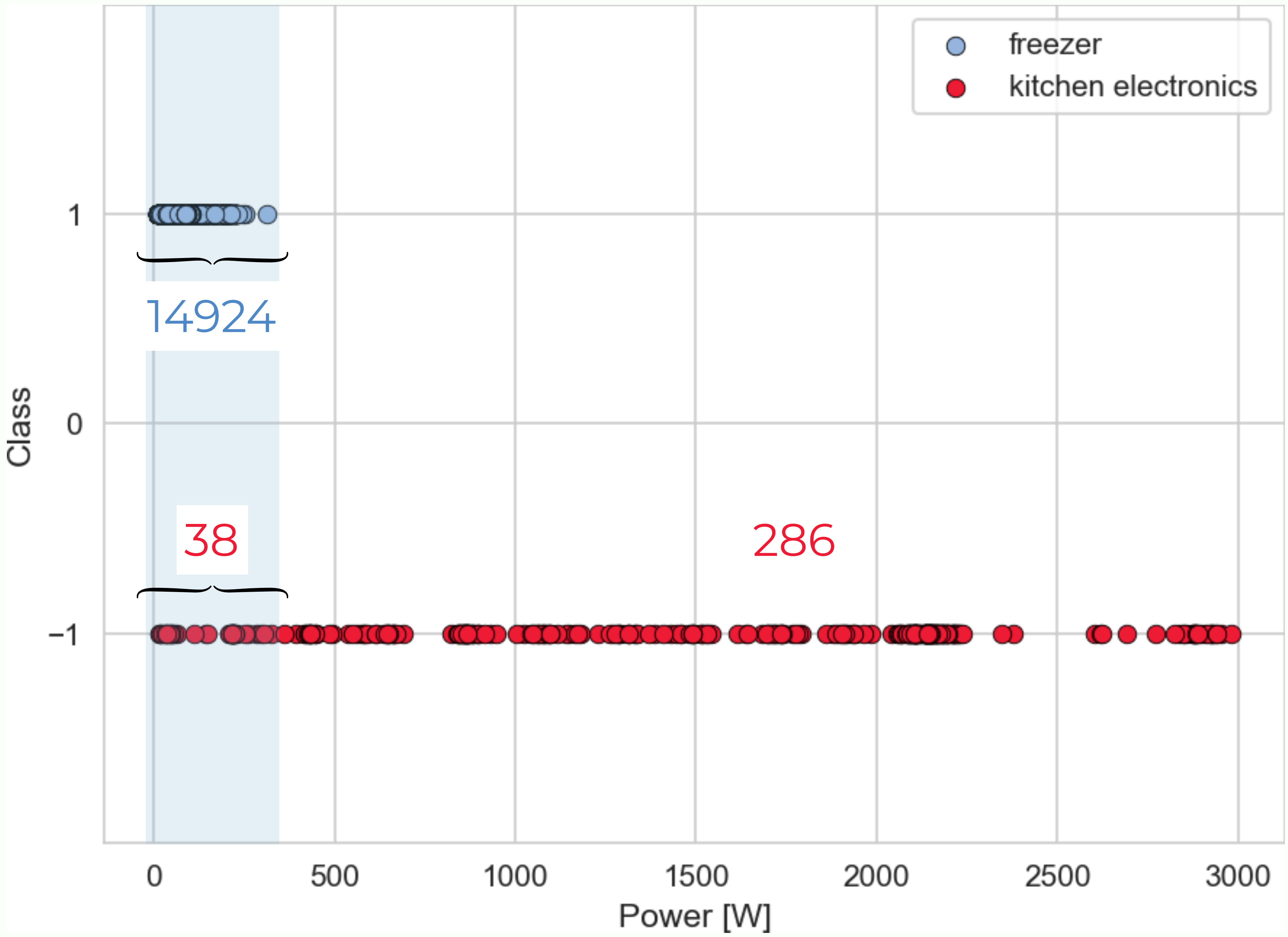


hard

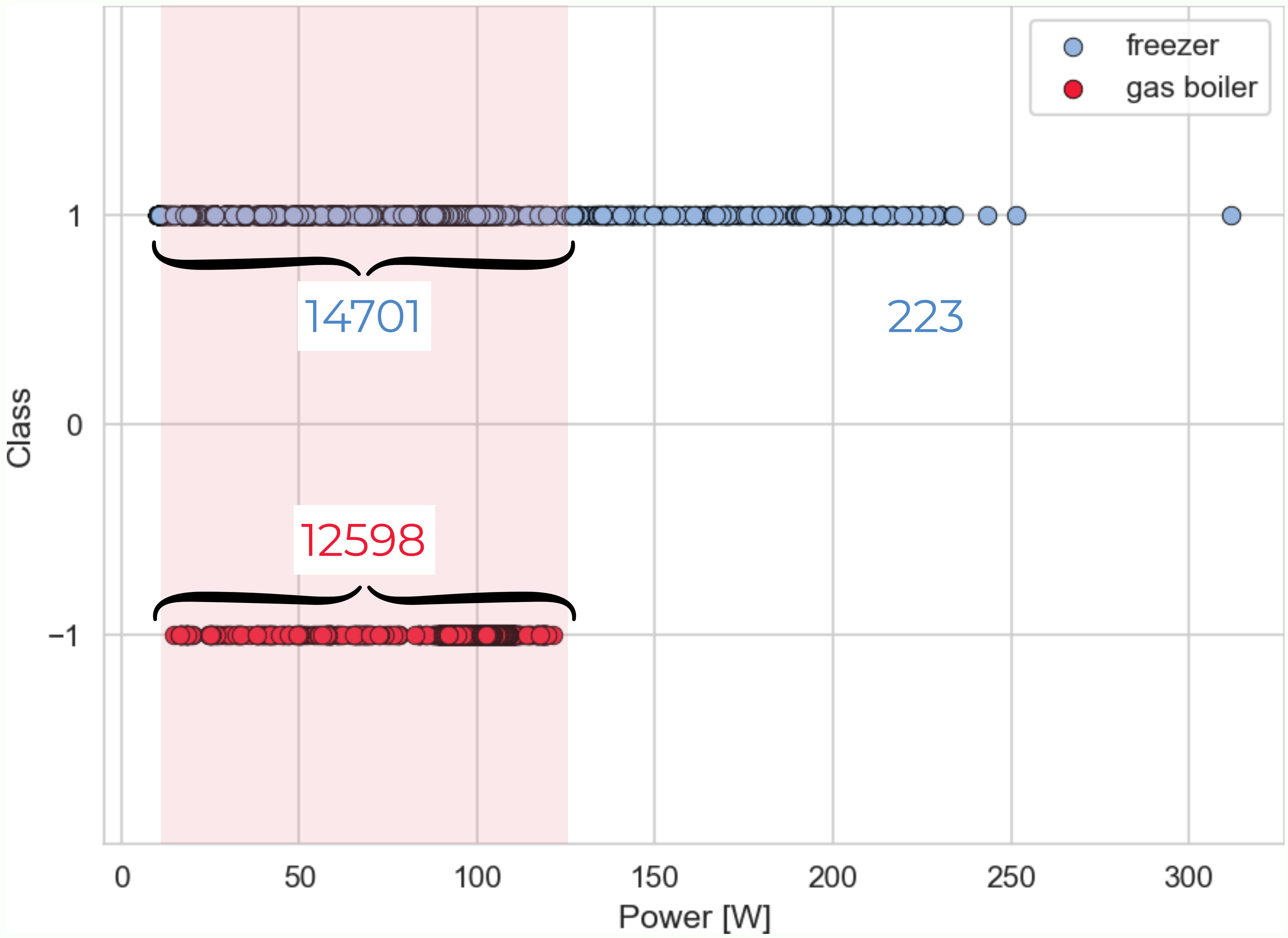


1D CASE: GEOMETRIC VIEW

easy

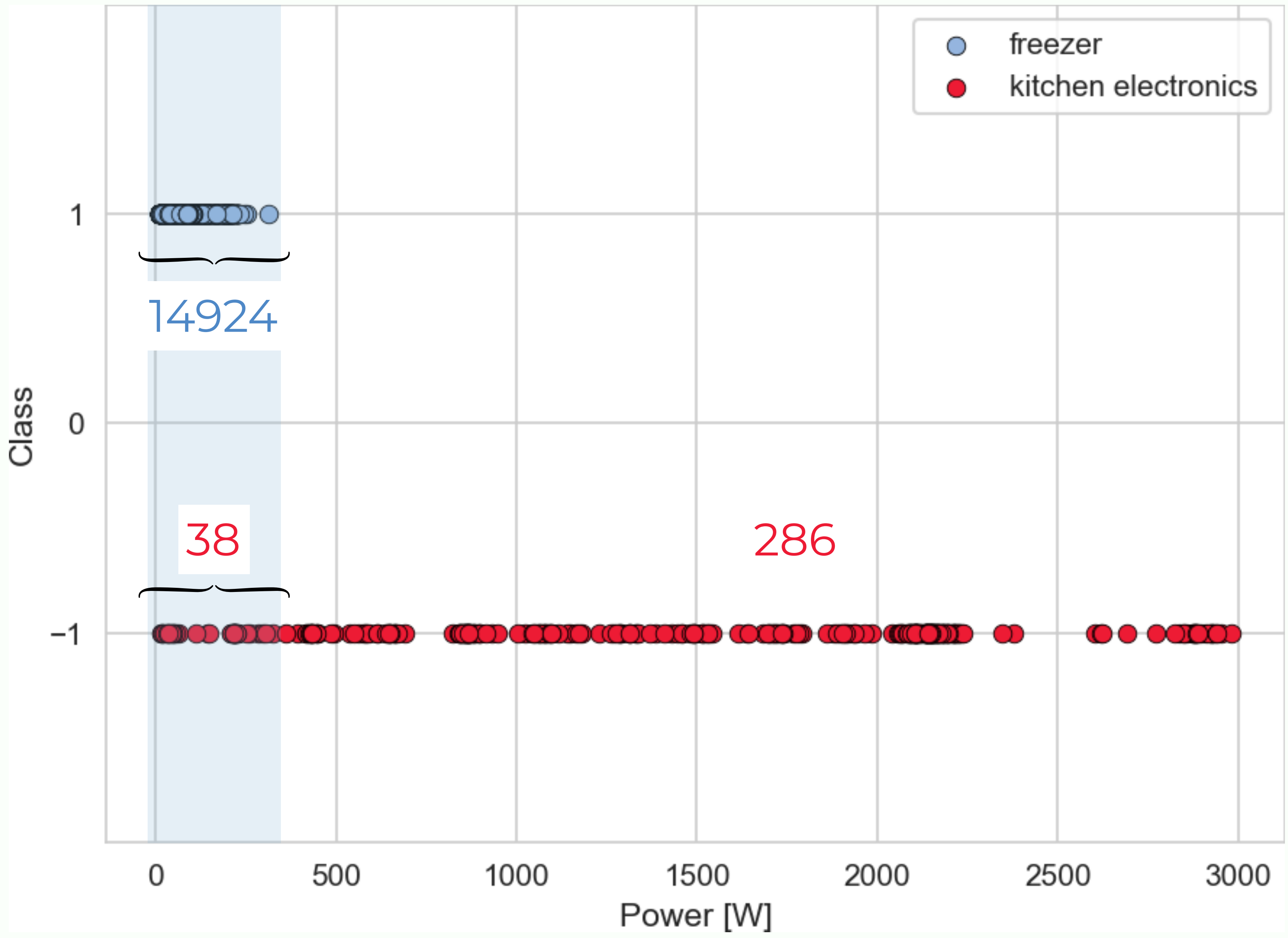


hard

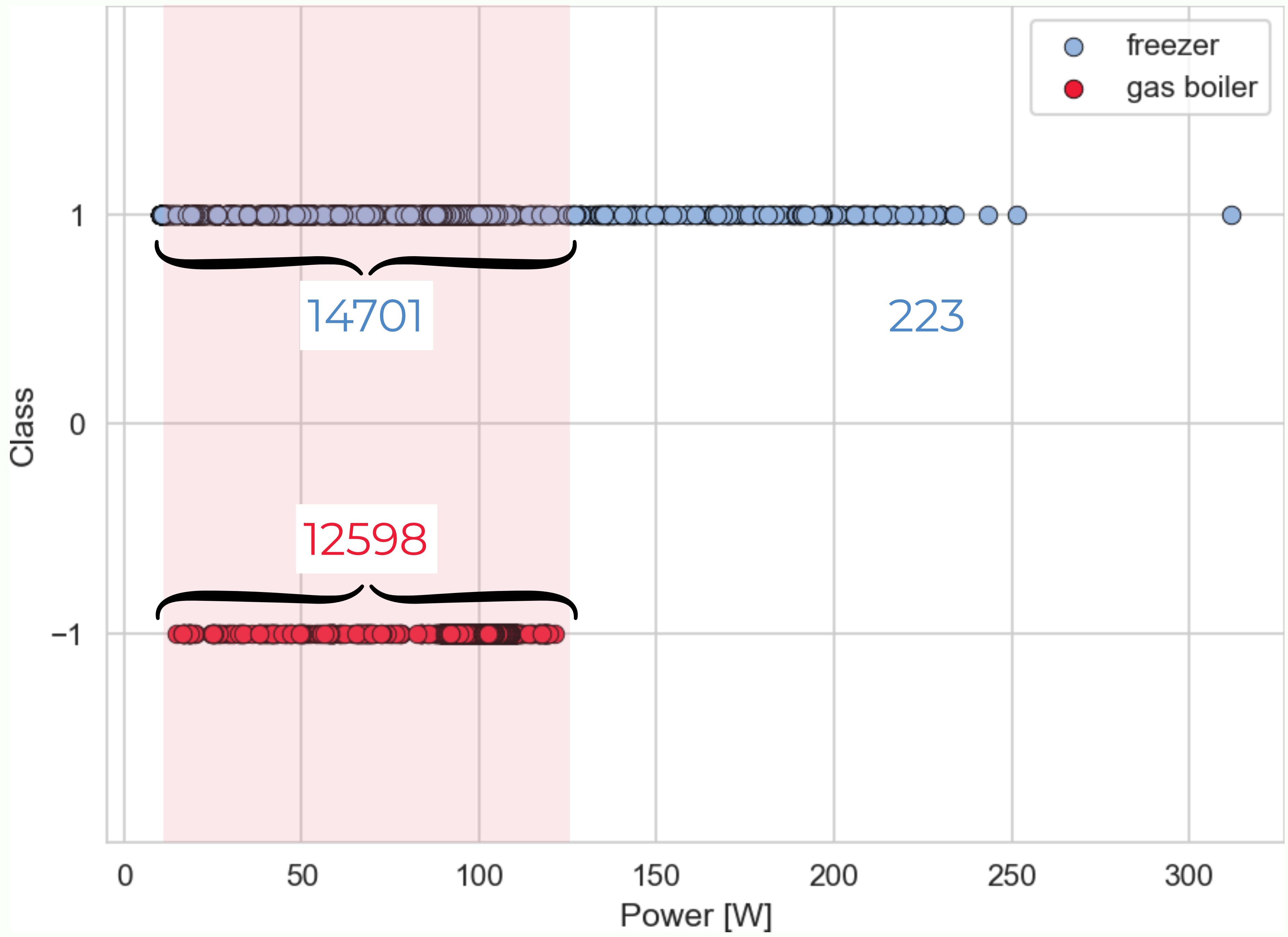


1D CASE: GEOMETRIC VIEW

easy



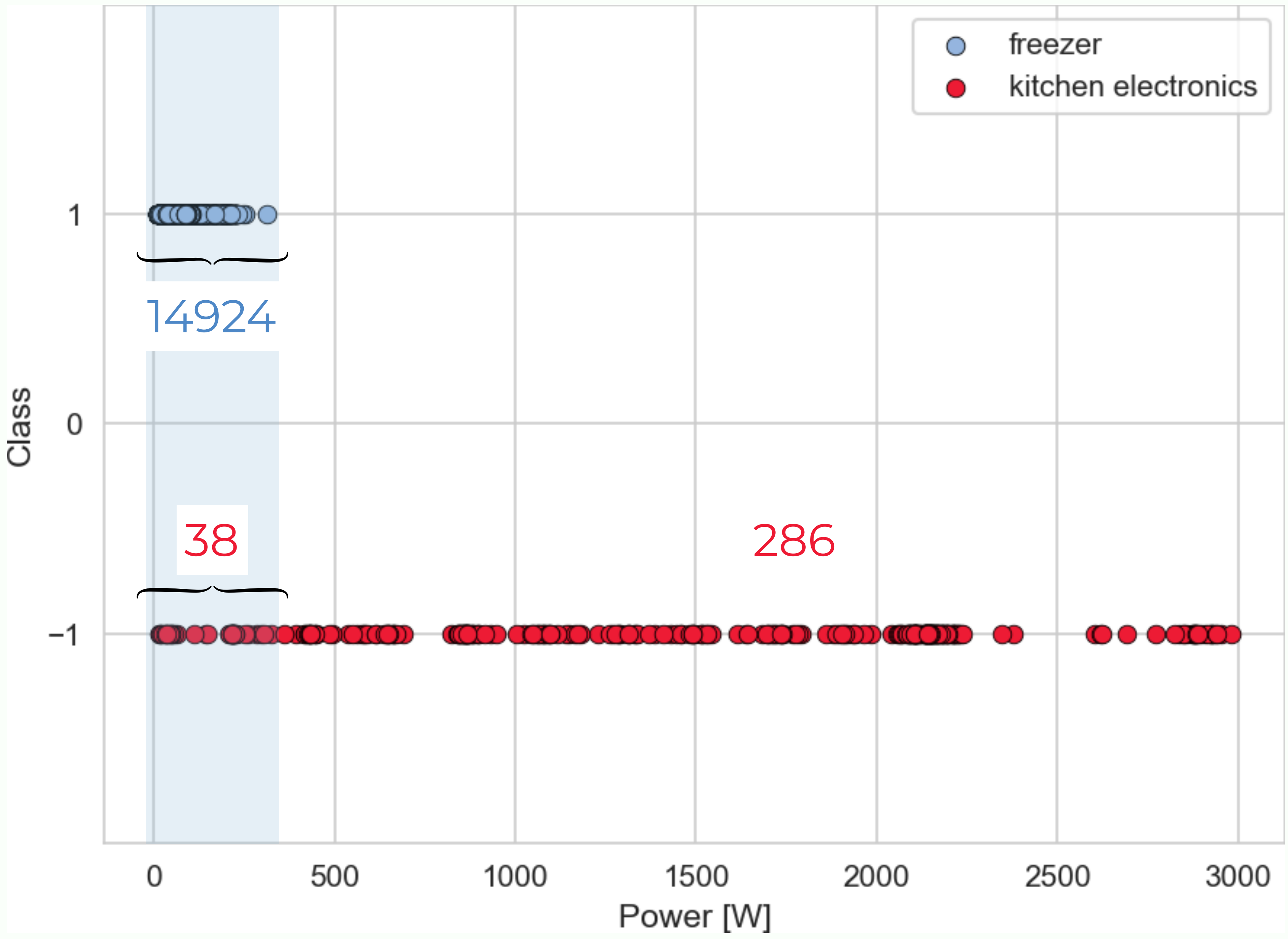
hard



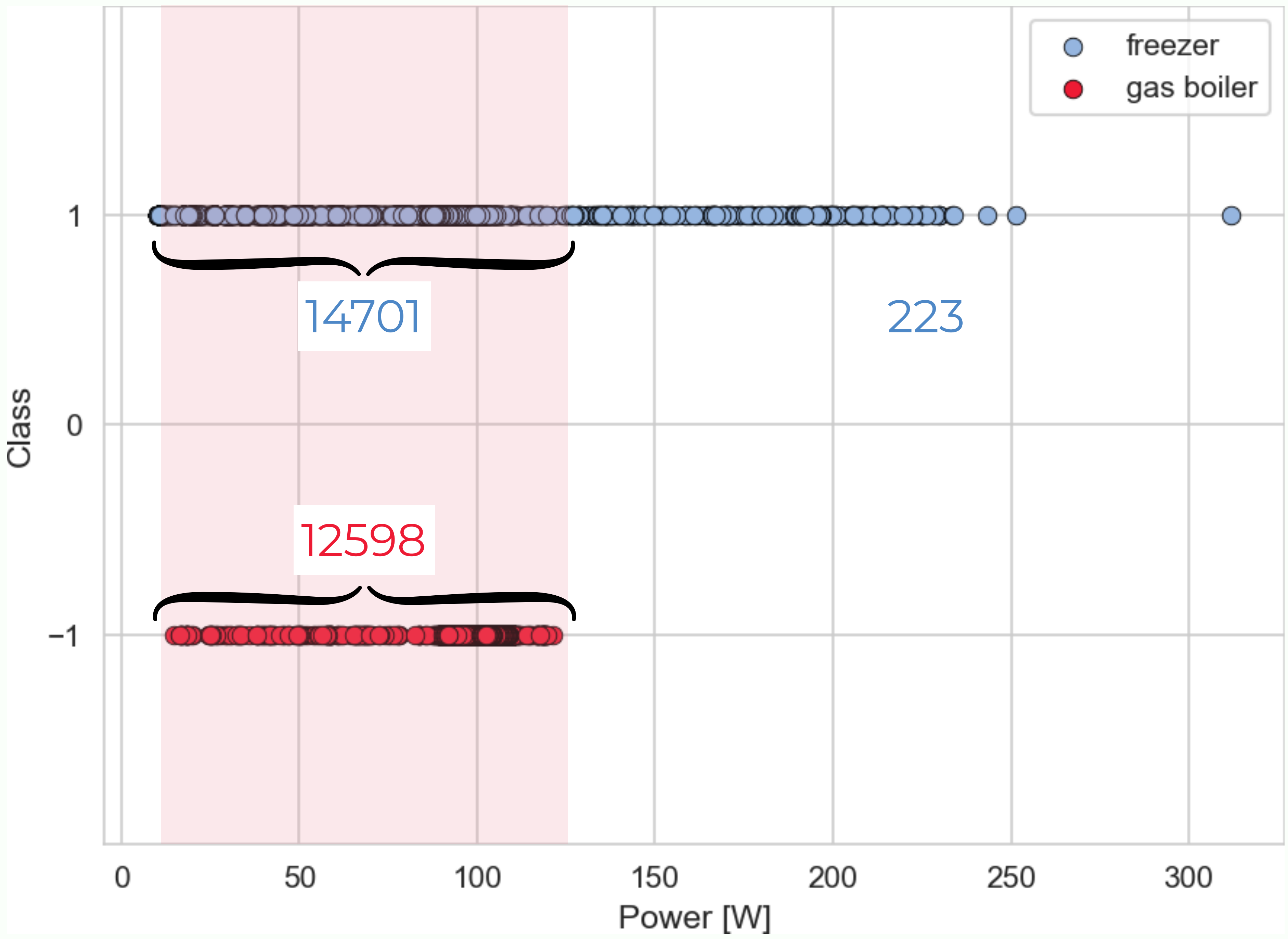
Conclusion?

1D CASE: GEOMETRIC VIEW

easy



hard



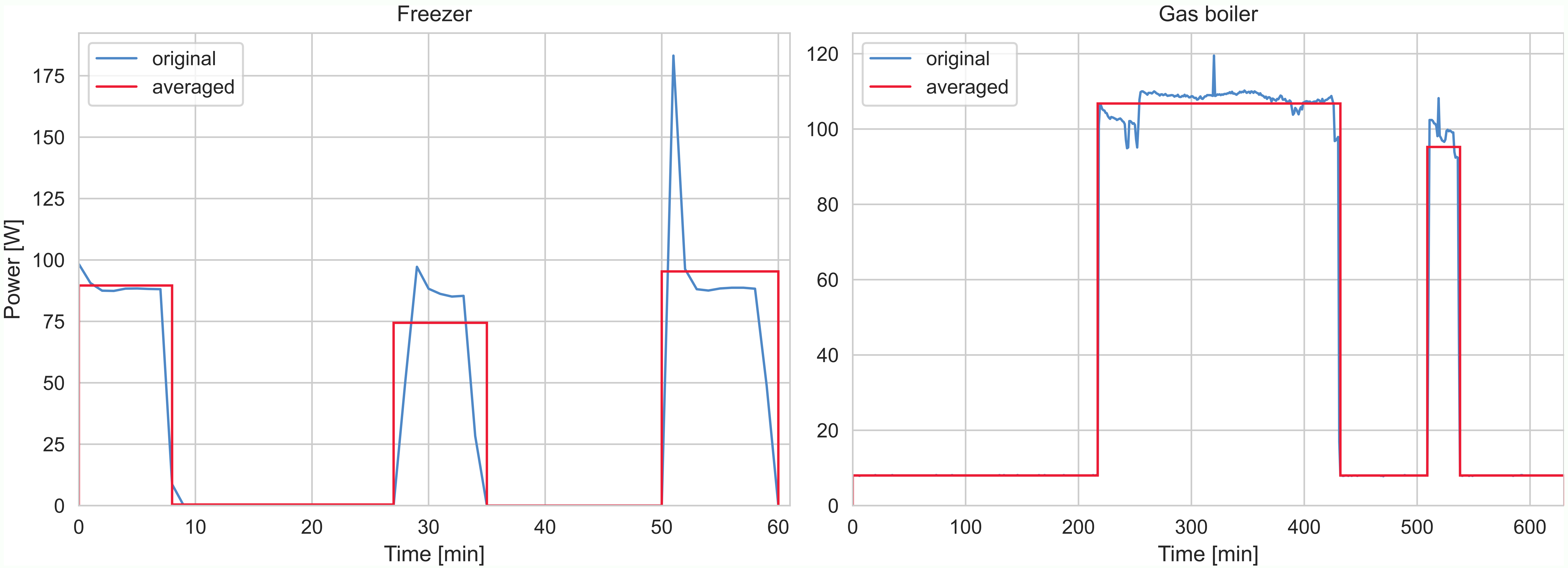
Conclusion?

We need more features :).

CONSTRUCTING NEW FEATURES

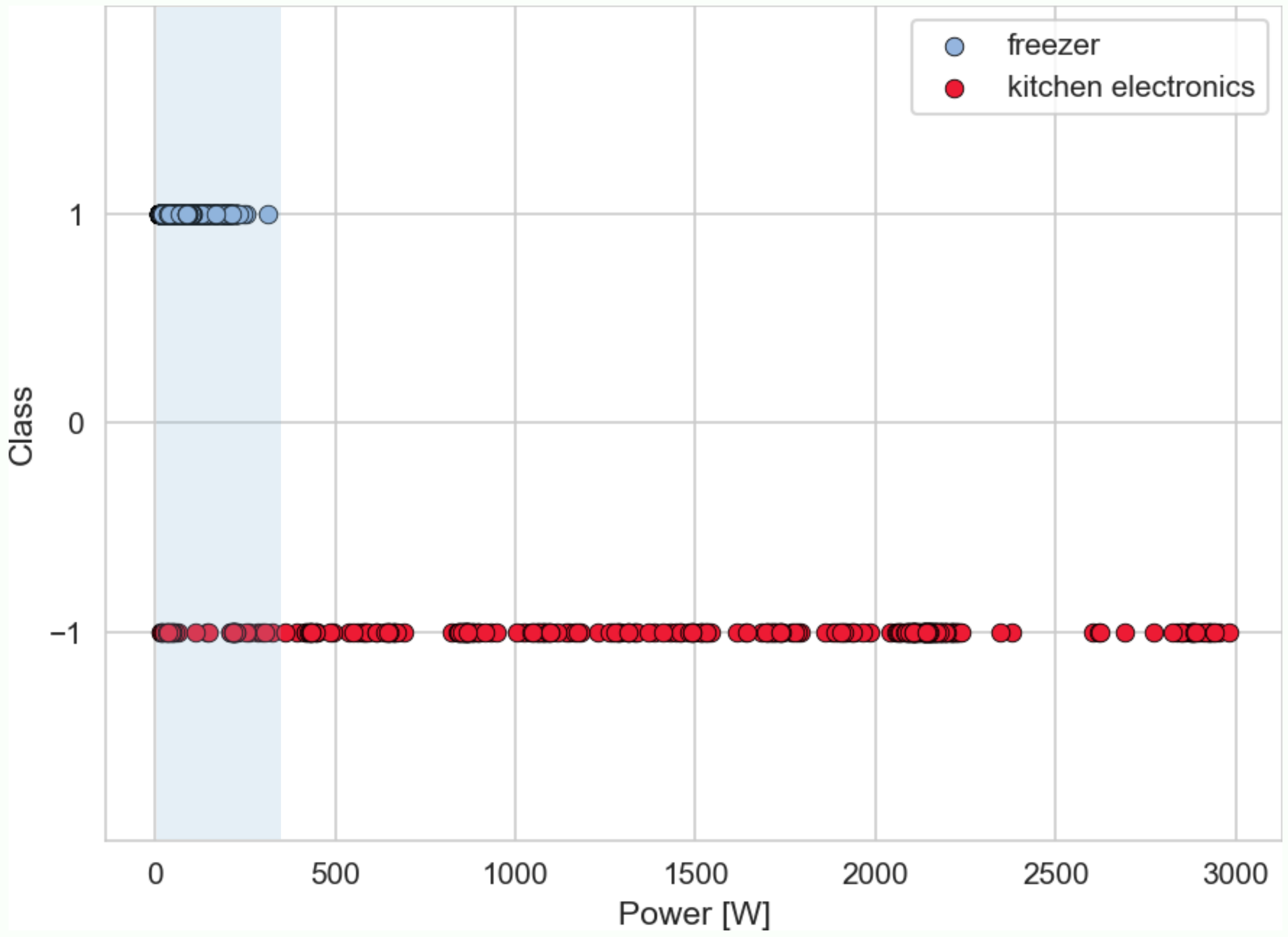
New features are:

- duration (d) of (in)activity periods,
- average power (P_{avg}) over these periods.

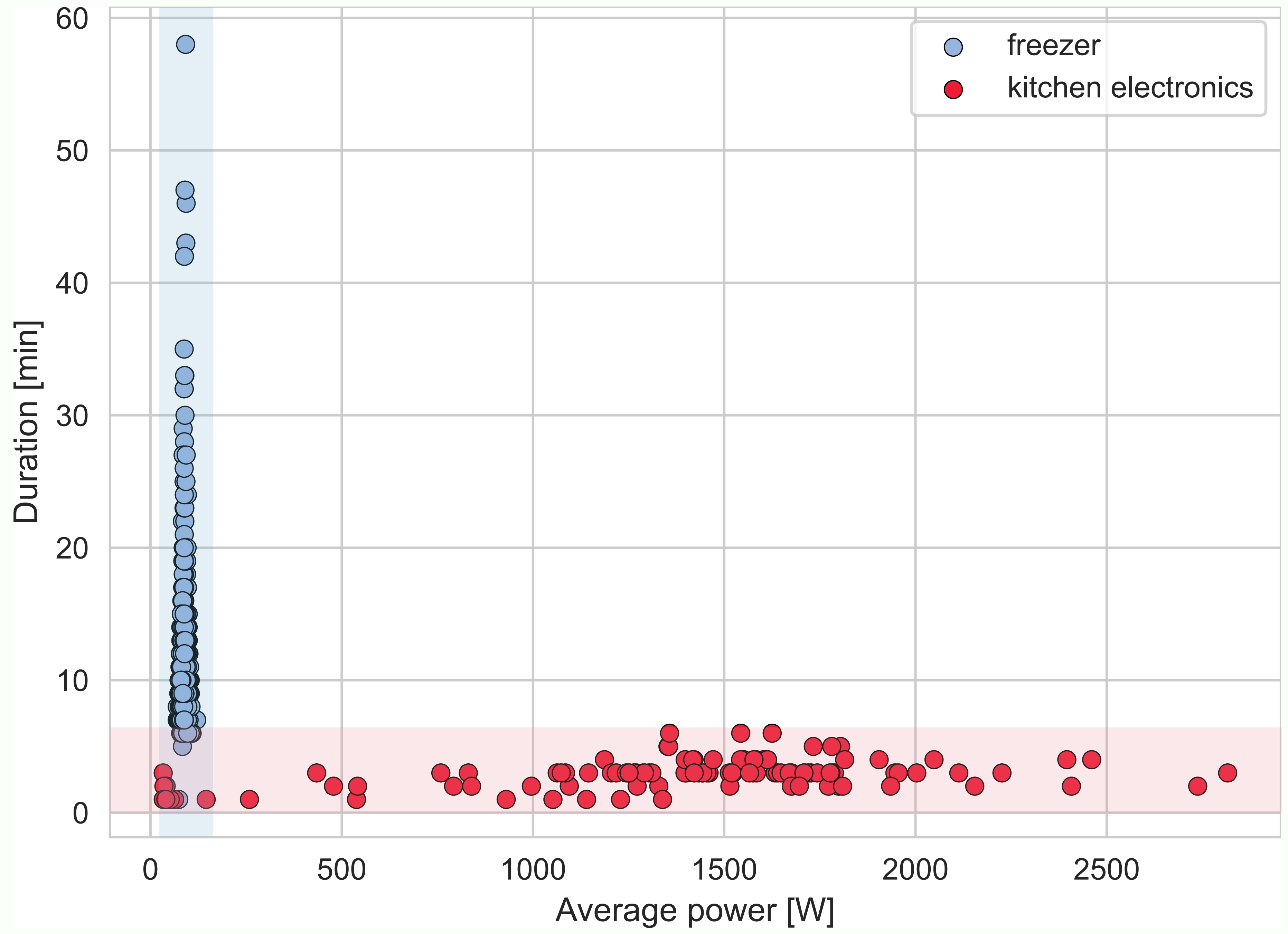


EASY CASE

in 1D: one feature

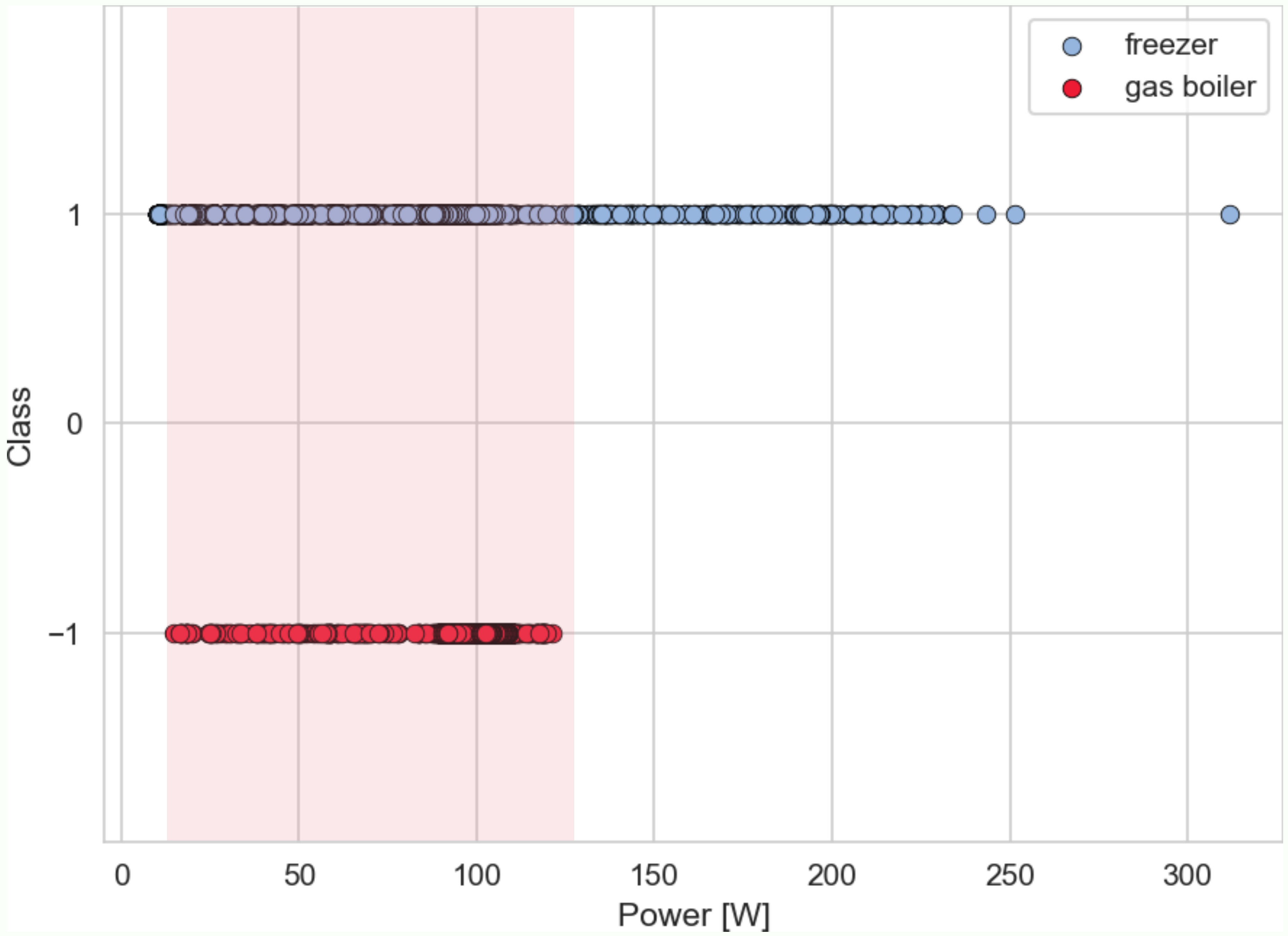


in 2D: two features

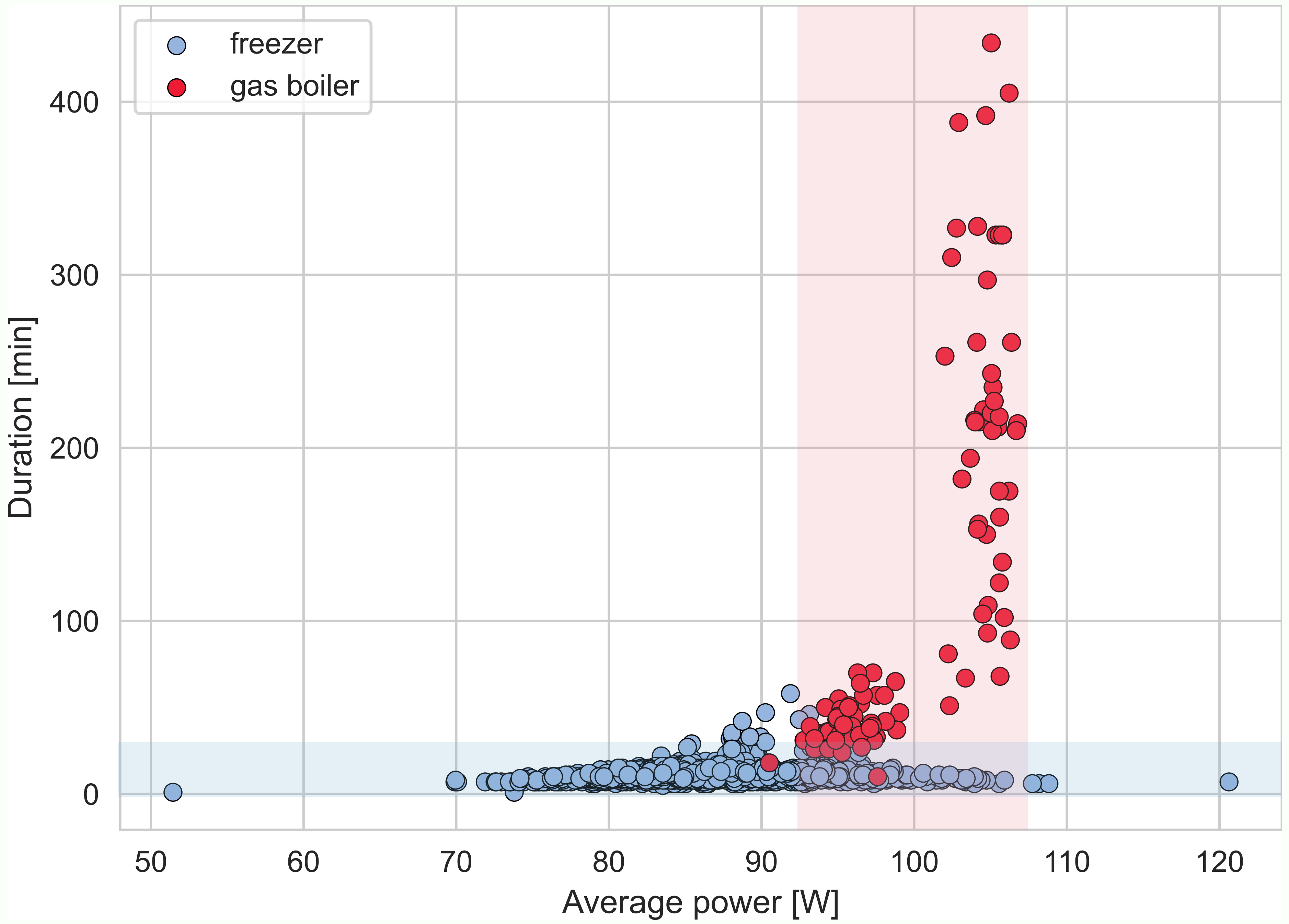


HARD CASE

in 1D: one feature

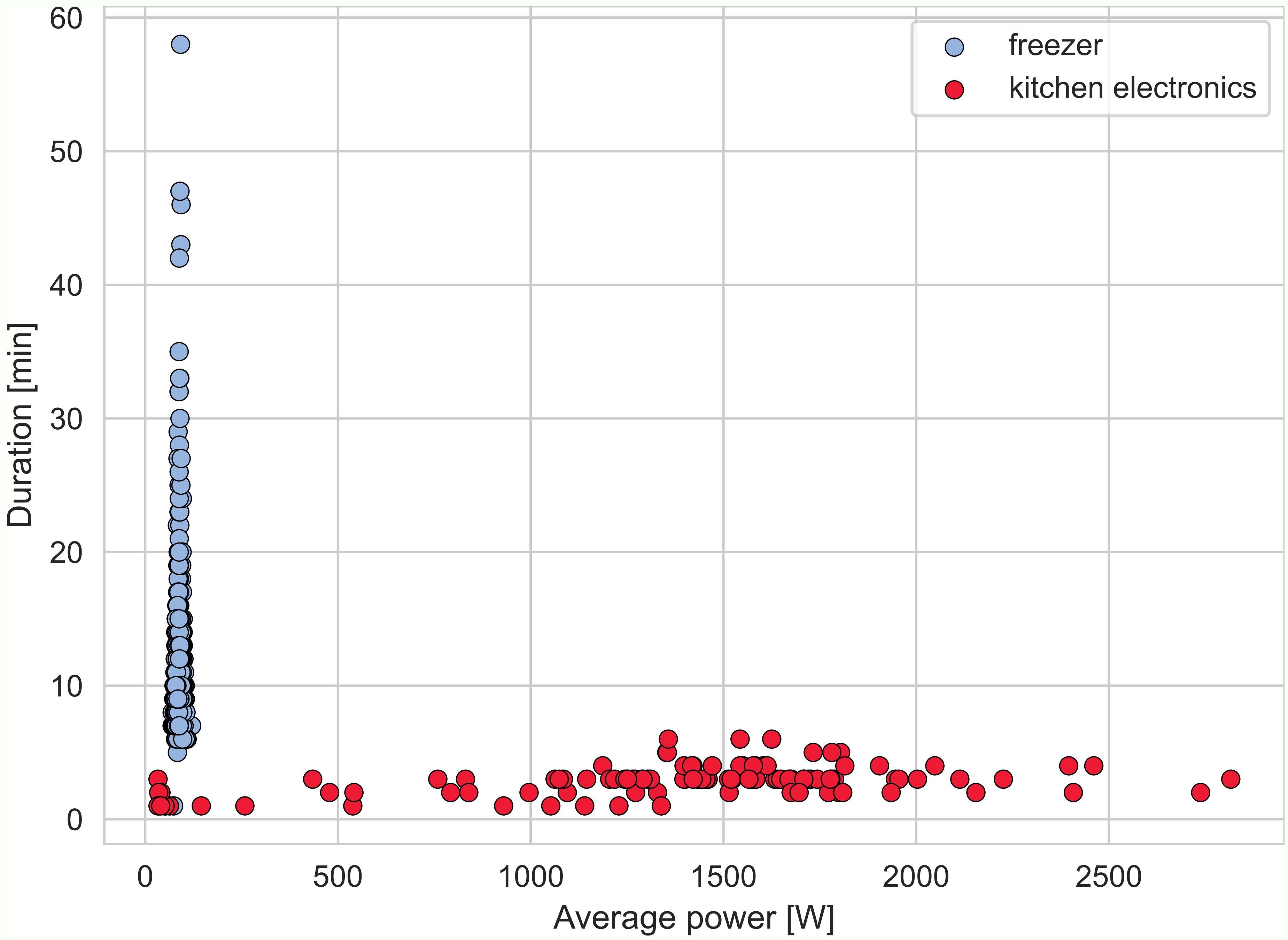


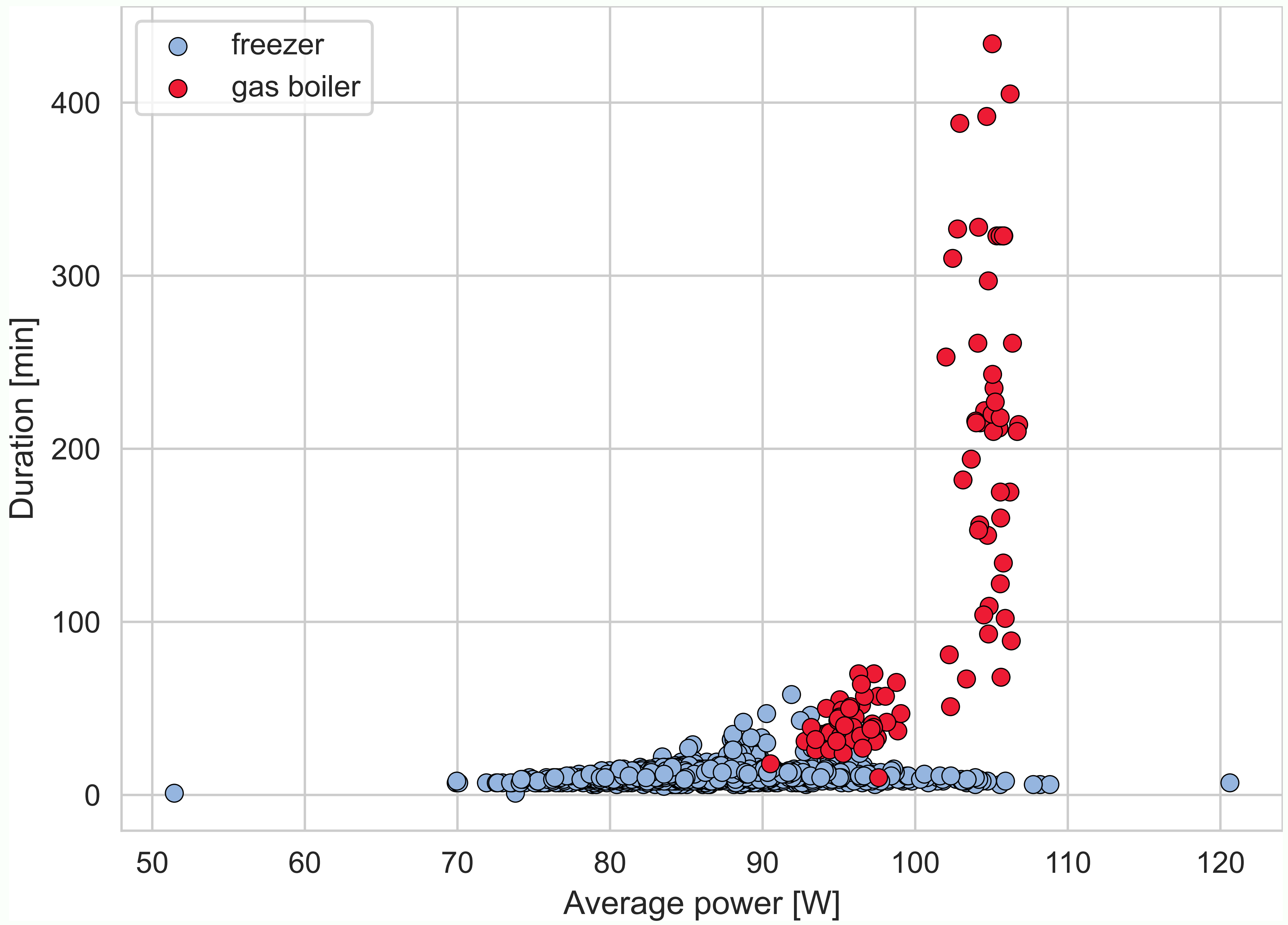
in 2D: two features



HAND-PICKED CLASSIFICATION

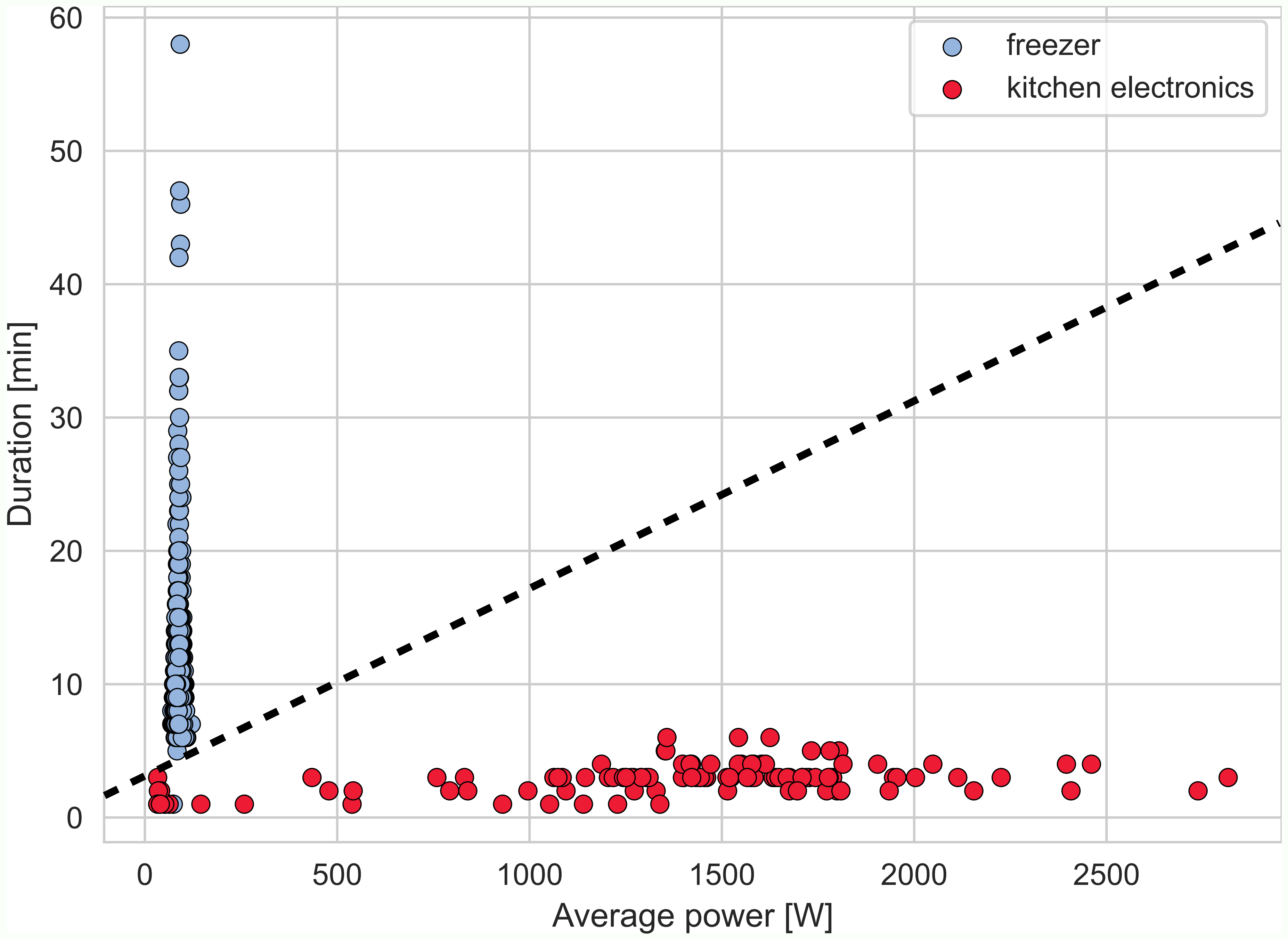
Linear classification = drawing a *line* separating classes.

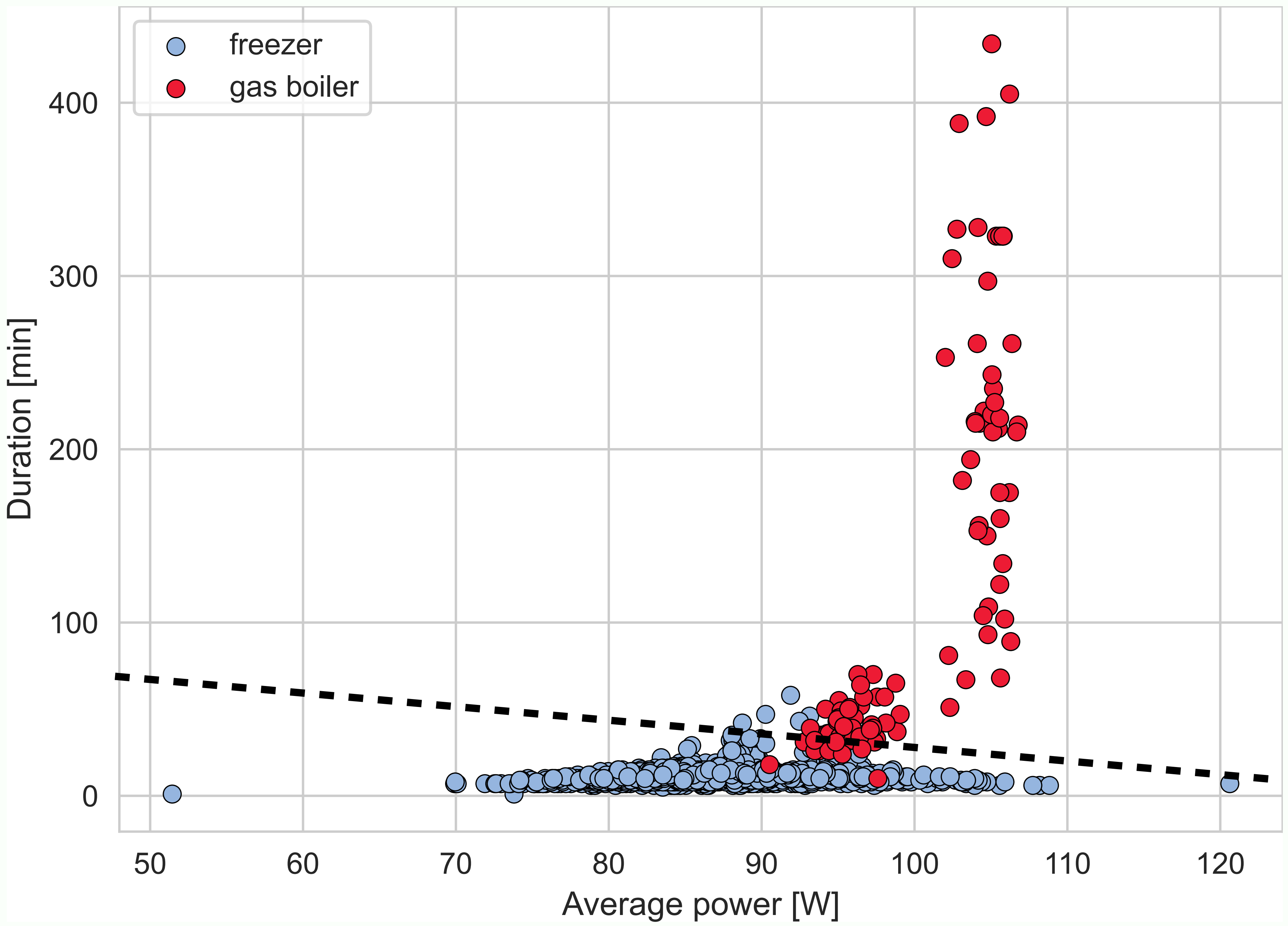




HAND-PICKED CLASSIFICATION

Linear classification = drawing a *line* separating classes.





FEATURE ENGINEERING

KEY COMPONENTS

Feature:

- creation (from data)
- scaling
- transformation
- extraction (from other features)
- selection

FEATURE CREATION

Feature creation is the process of generating new features based on domain knowledge or by observing patterns in the data.

Types of feature creation:

- combination/aggregation
- domain-specific
- data-driven
- synthetic

EXAMPLES

From “Date” we may get (temporal features):

- Hour of day, day of week, weekday/working day, etc.

Electricity demand data (in MWh)

Date (Estonia time)	Demand
01.01.2023 00:00	798,2
01.01.2023 01:00	793,4
01.01.2023 02:00	776,5
01.01.2023 03:00	757
01.01.2023 04:00	743,7
01.01.2023 05:00	737,6
01.01.2023 06:00	749,4
...	...

From “Demand” we may get:

- Magnitude, high/low (trend), peak/bottom demand, etc.

EXAMPLES (2)

Electricity production, Estonia, 2024

Month	Solar [MWh]
Aug	121549
Jul	139590
Jun	162682
May	183217
Apr	82439
Mar	64077
Feb	19145
Jan	4385

[Accessed September 30, 2025], https://stat.ee/en/find-statistics/statistics-theme/population

EXAMPLES (2)

Electricity production, Estonia, 2024

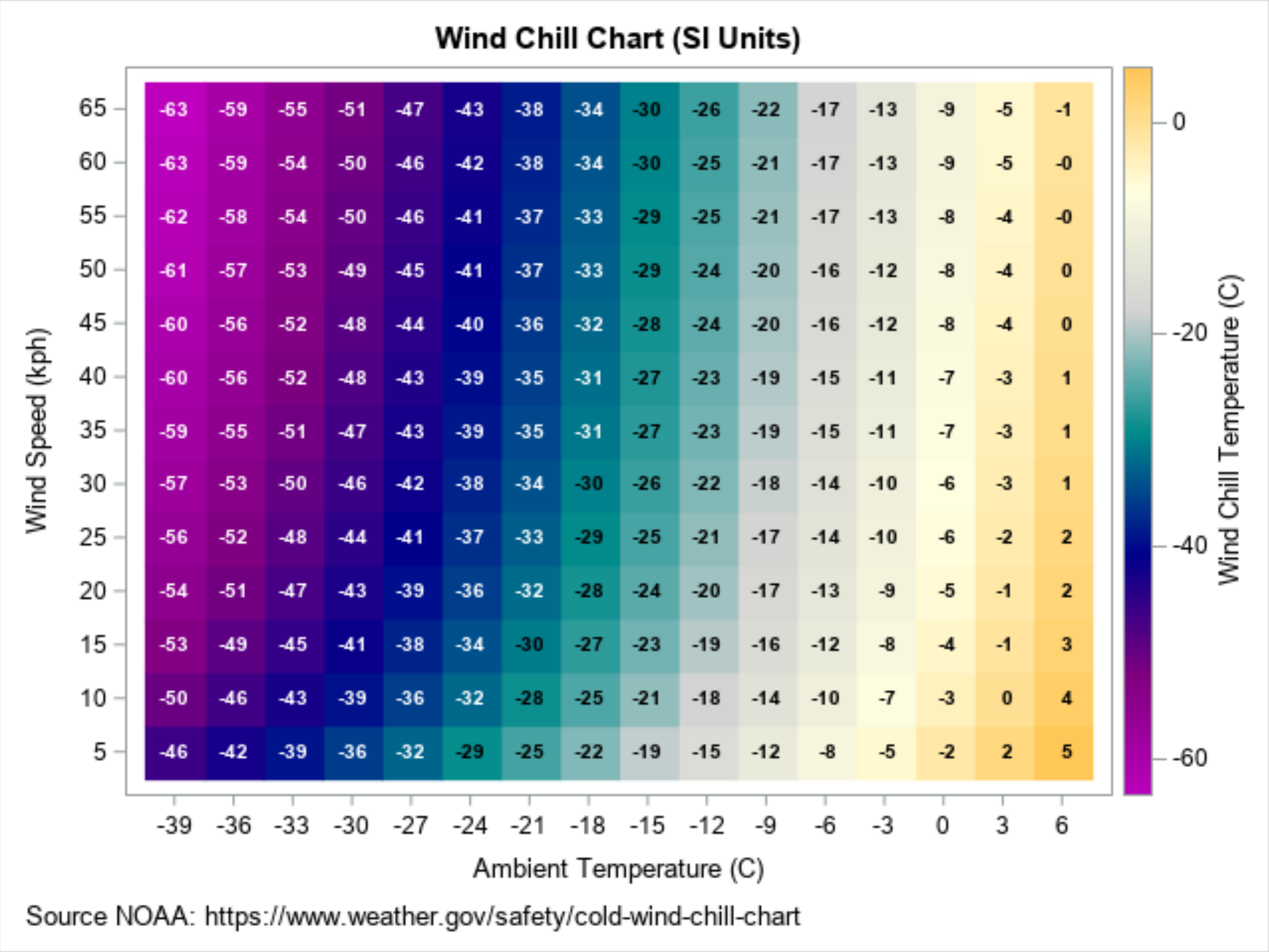
Month	Solar [MWh]
Aug	121549
Jul	139590
Jun	162682
May	183217
Apr	82439
Mar	64077
Feb	19145
Jan	4385

Estonia population (2024) was 1.374 mln

Solar per capita [kWh/popul. size]
88.46
101.59
118.4
133.34
59.99
46.64
13.93
3.19

[Accessed September 30, 2025], https://stat.ee/en/find-statistics/statistics-theme/population

EXAMPLES (3)



$$\begin{aligned}\mathrm{WCT}_{\mathrm{SI}} = 13.12 + 0.6215T_C \\ -11.37V_{kph}^{0.16} + 0.3965T_C V_{kph}^{0.16}\end{aligned}$$

[Accessed September 14, 2025], https://blogs.sas.com/content/iml/2021/02/24/wind-chill-chart.html

FEATURE SCALING

Feature scaling is the process of transforming the features so that they lie within a comparable scale or range.

- Important for distance-based algorithms (SVM or kNN)
- Improves converge speed (gradient descent)
- Improves accuracy
- Features have equal impact

SELECTED METHODS

Min-max normalisation to $[0,1]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

Min-max normalisation to $[a,b]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}(b - a) + a.$$

SELECTED METHODS

Min-max normalisation to $[0,1]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

Min-max normalisation to $[a, b]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}(b - a) + a.$$

Mean normalisation:

$$\tilde{x} = \frac{x - \mu}{\max(x) - \min(x)}.$$

μ is the mean

SELECTED METHODS

Min-max normalisation to $[0,1]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

Min-max normalisation to $[a, b]$:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}(b - a) + a.$$

Mean normalisation:

$$\tilde{x} = \frac{x - \mu}{\max(x) - \min(x)}.$$

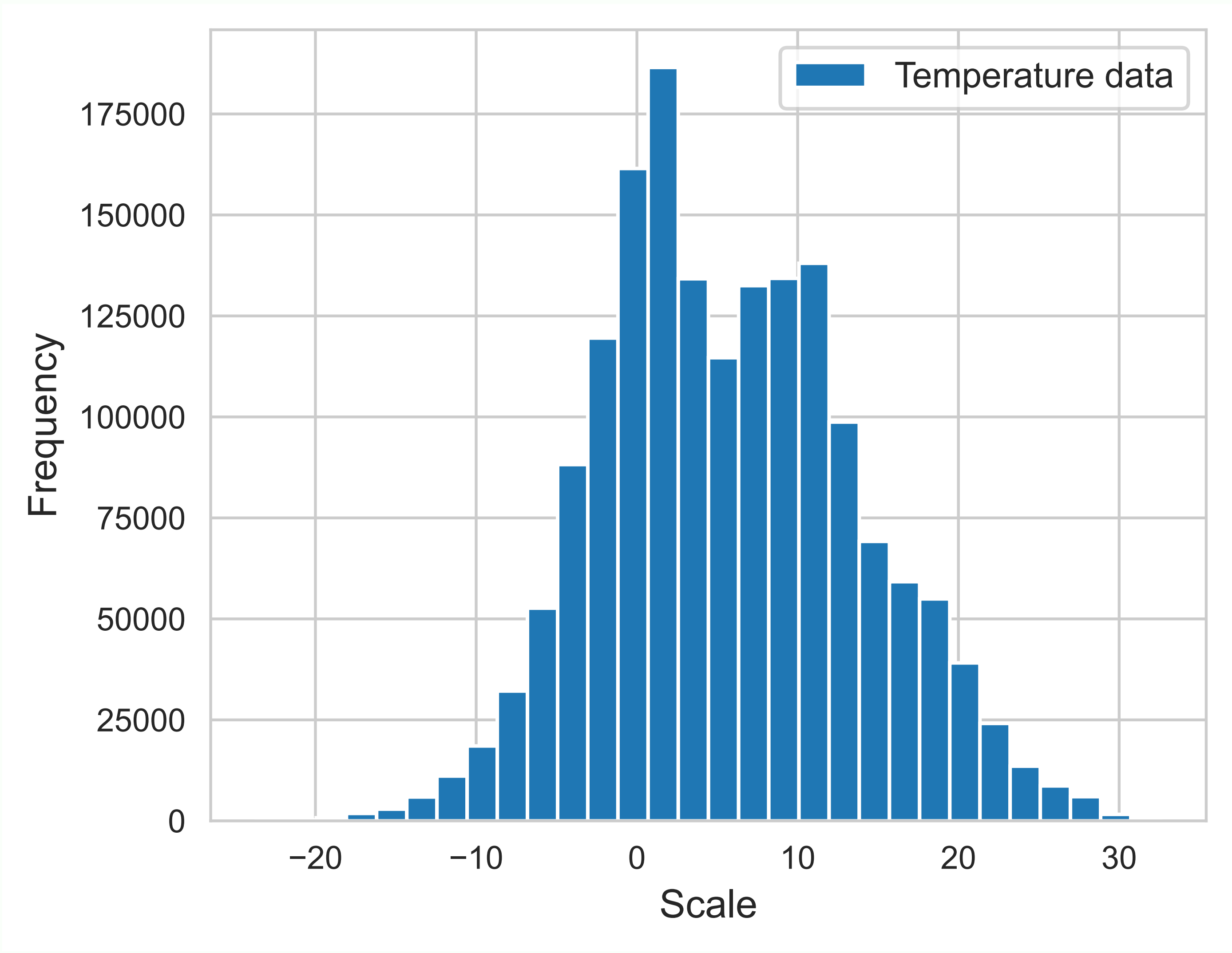
μ is the mean

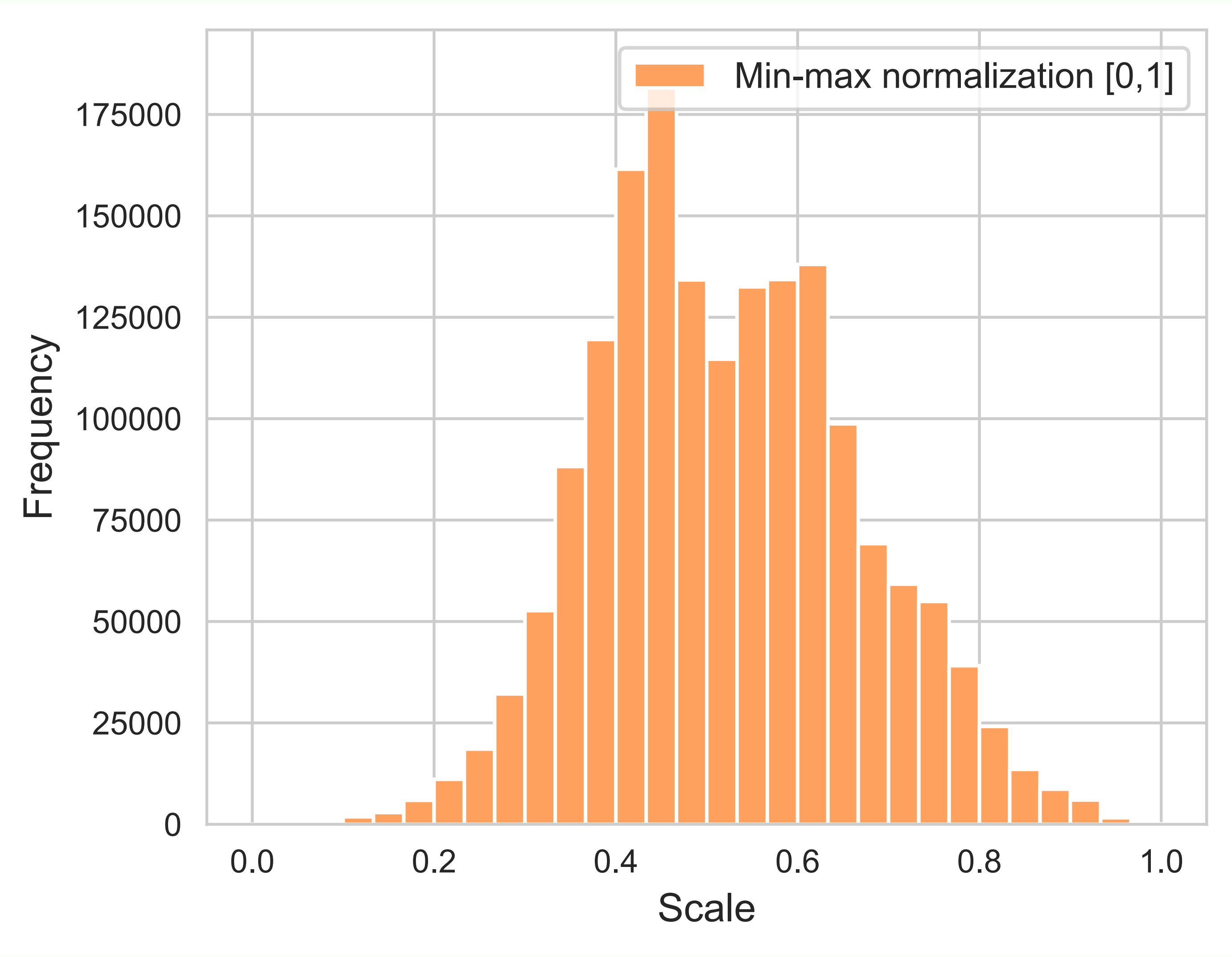
Standardisation (or z-score normalisation):

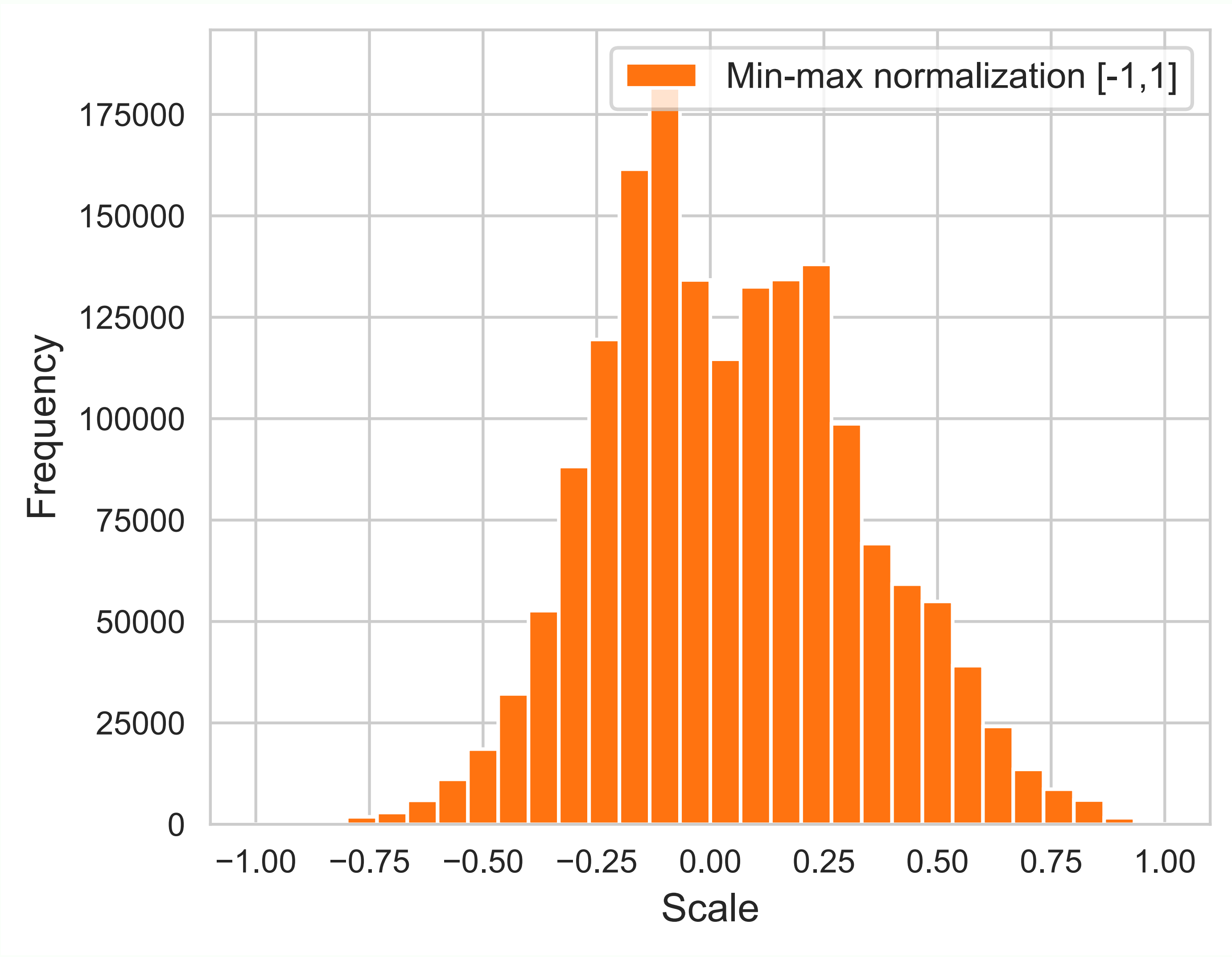
$$\tilde{x} = \frac{x - \mu}{\sigma}.$$

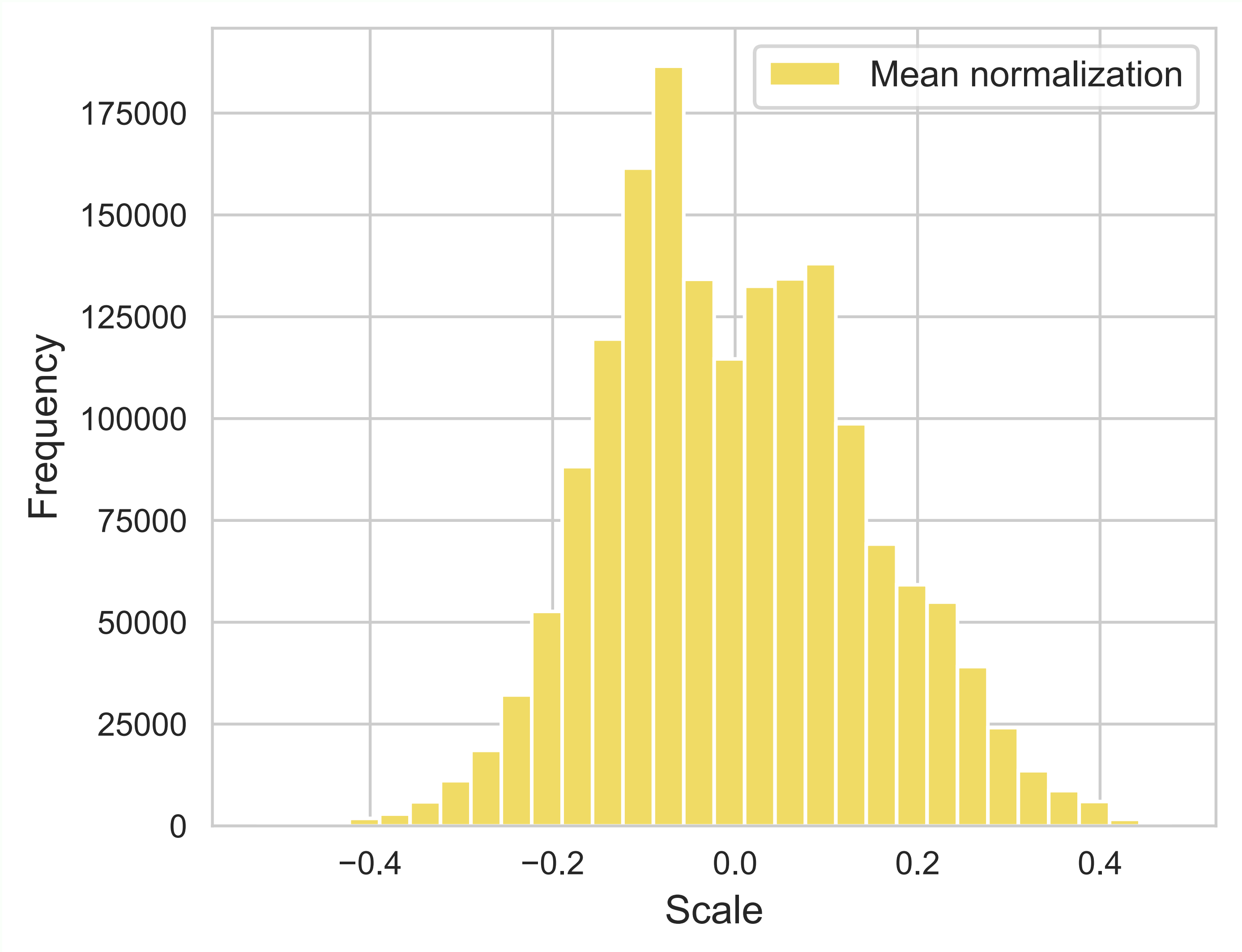
σ is the standard deviation

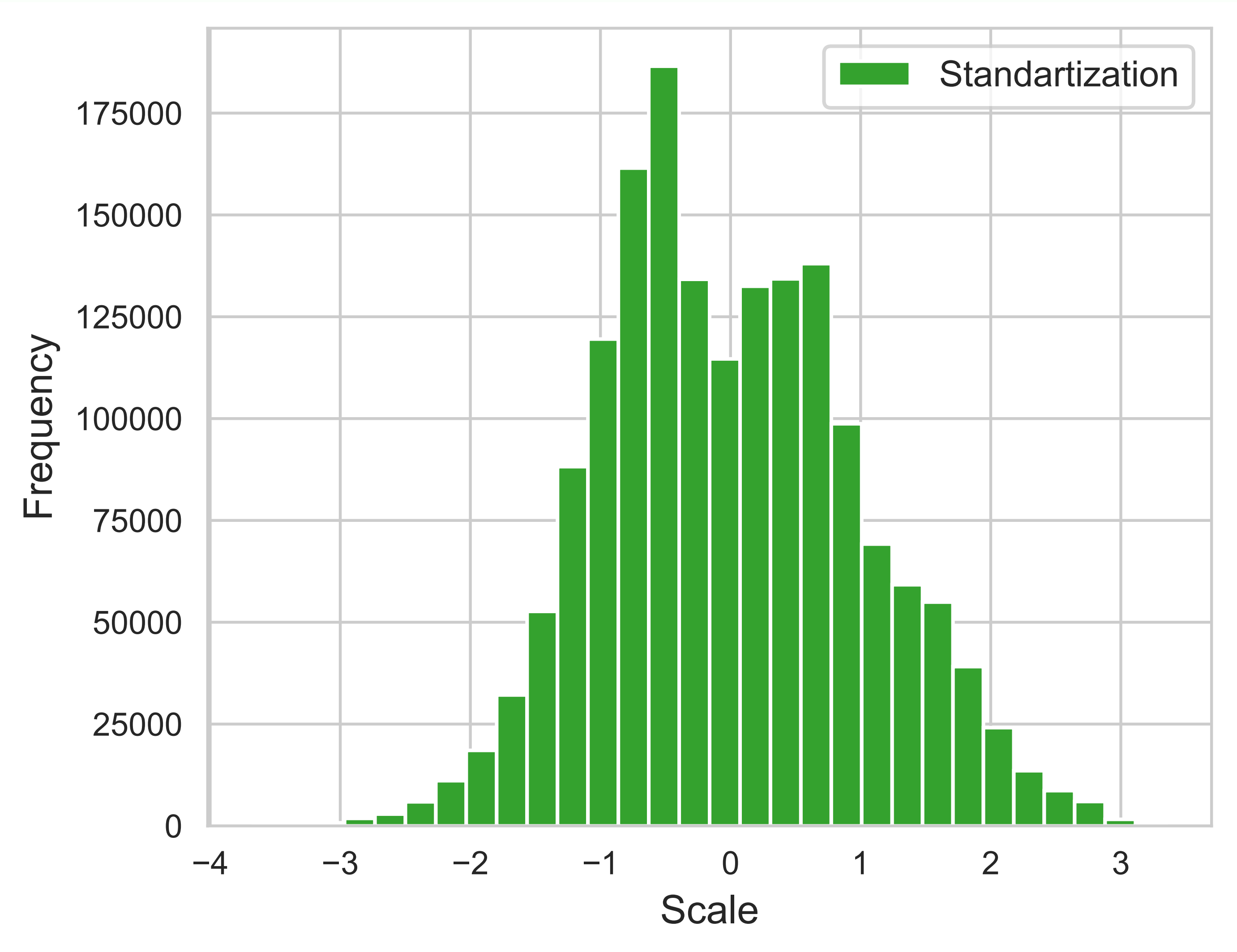
EXAMPLE





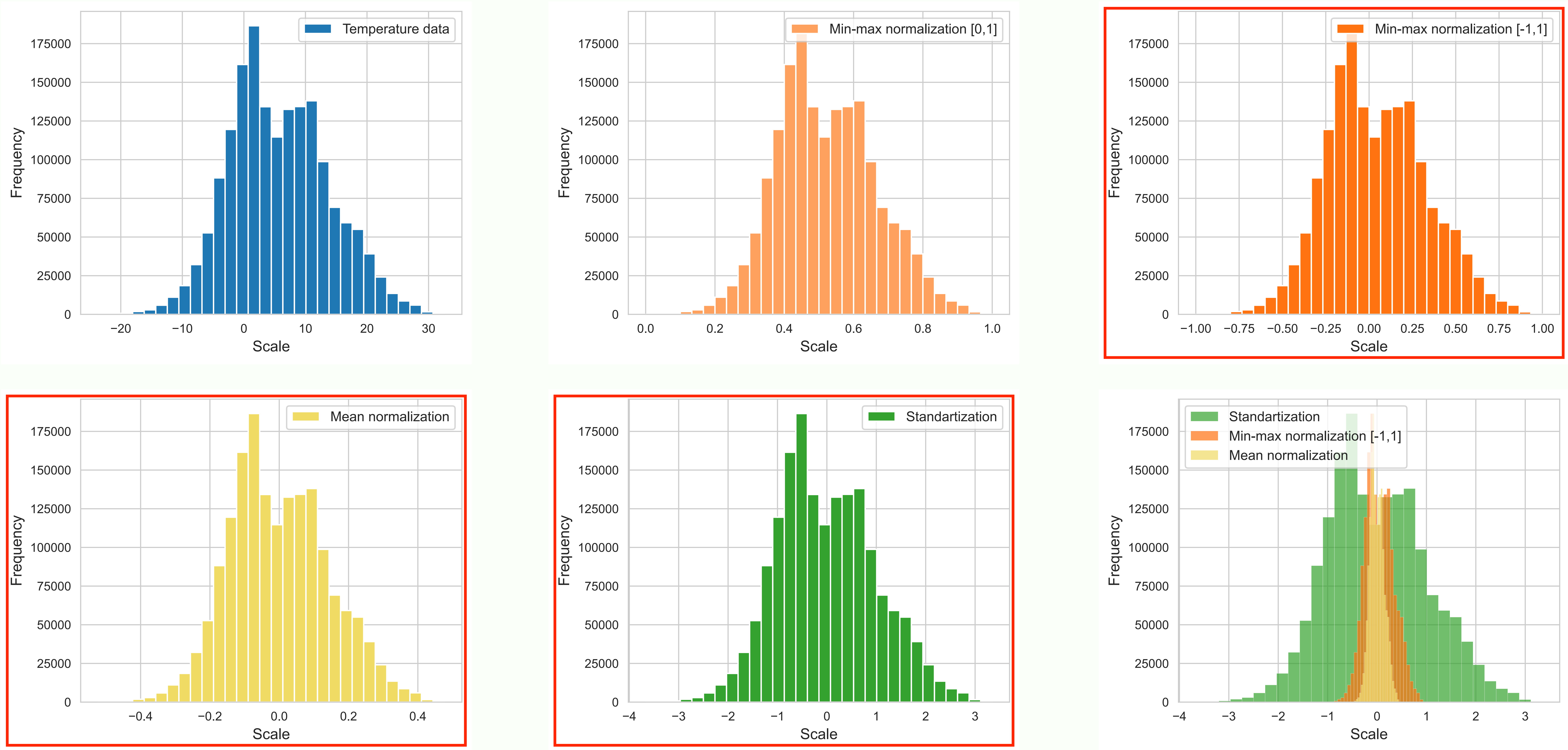






[Accessed September 13, 2025], https://www.kaggle.com/competitions/predict-energy-behavior-of-prosumers

EXAMPLE

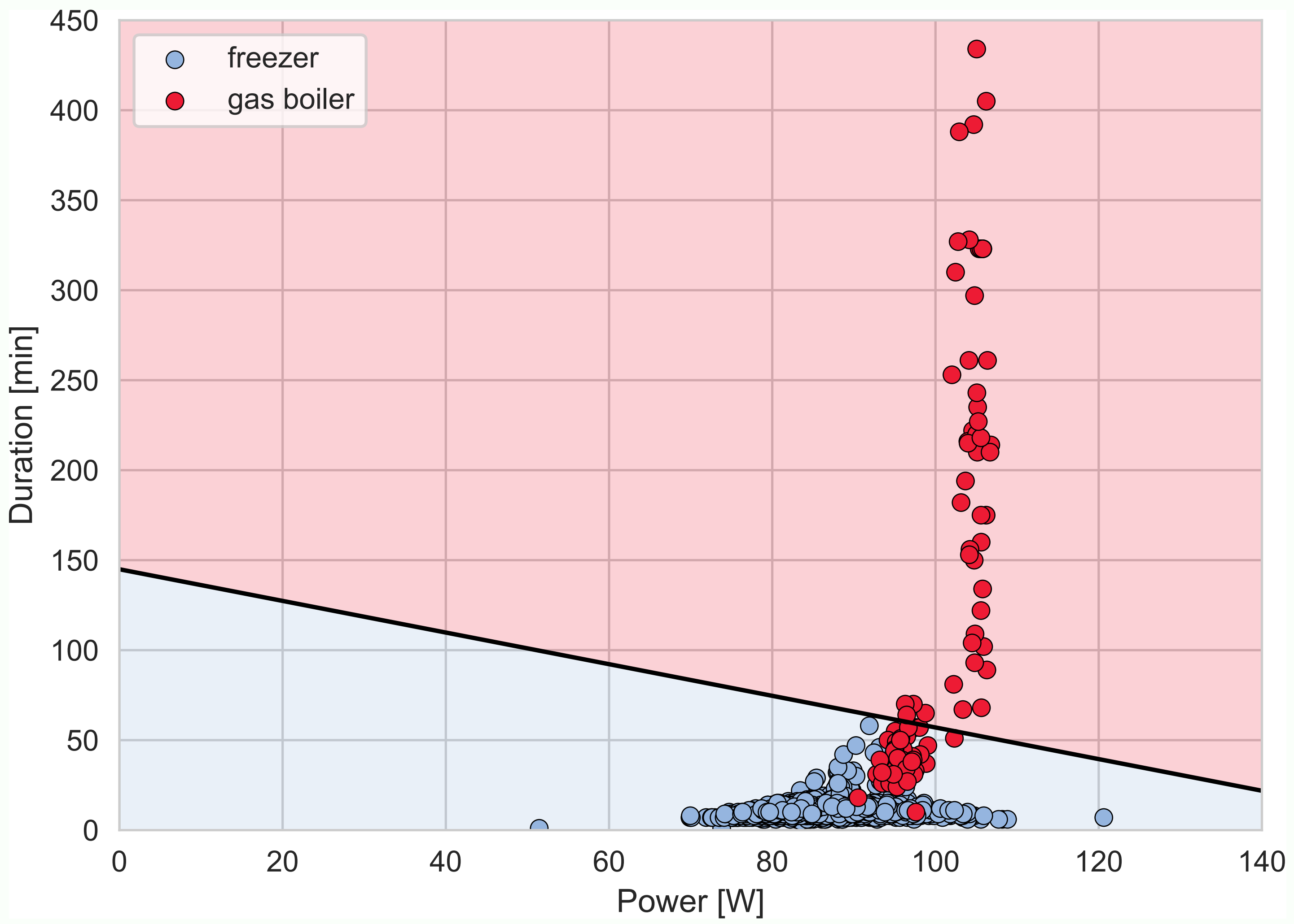


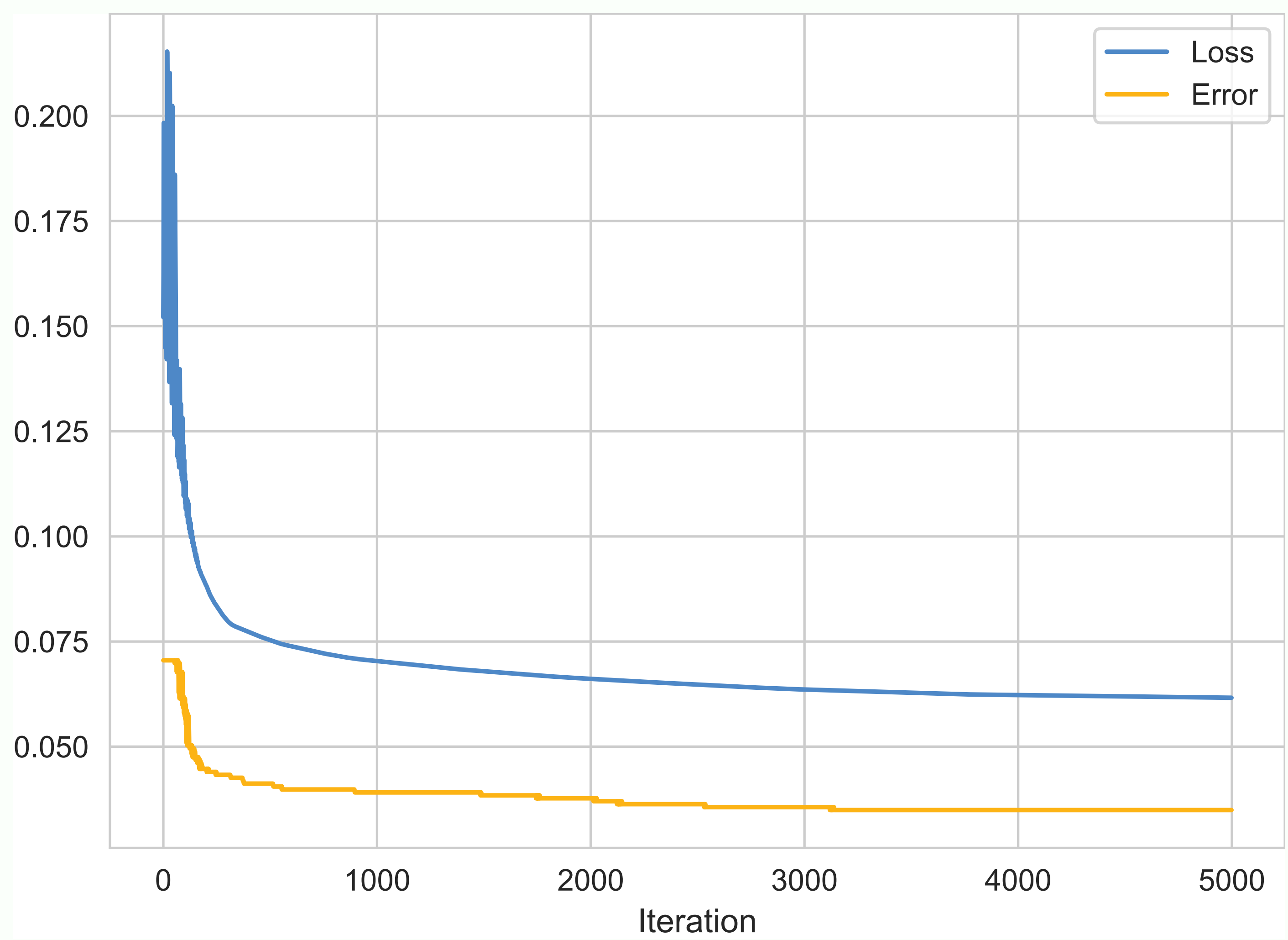
[Accessed September 13, 2025], https://www.kaggle.com/competitions/predict-energy-behavior-of-prosumers

What can go right or … wrong?

PRACTICAL OBSERVATIONS

BEFORE (min-max)



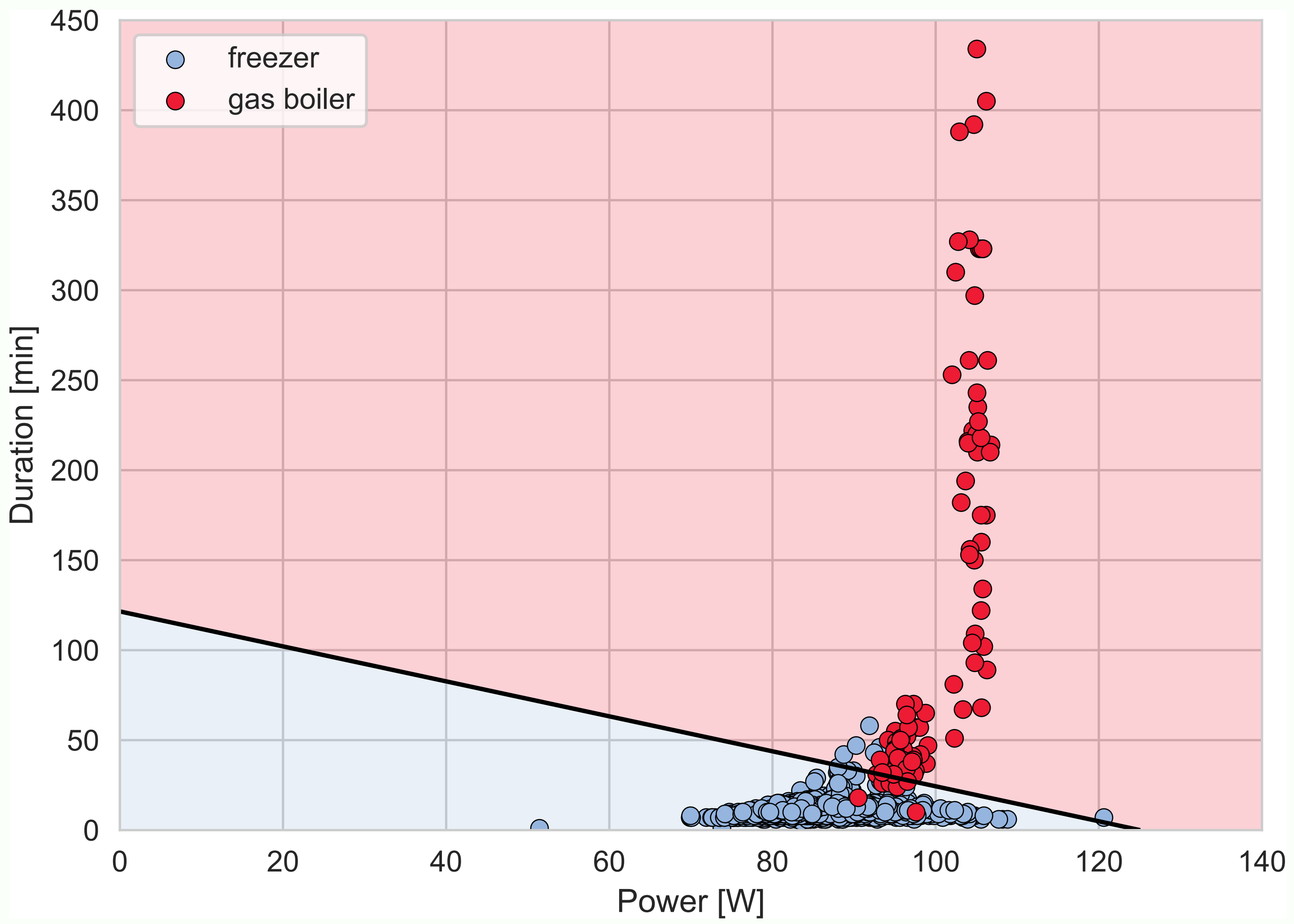


$\text{iter} = 5000$

$\lambda = 10^{-4}$

$\alpha = 1$

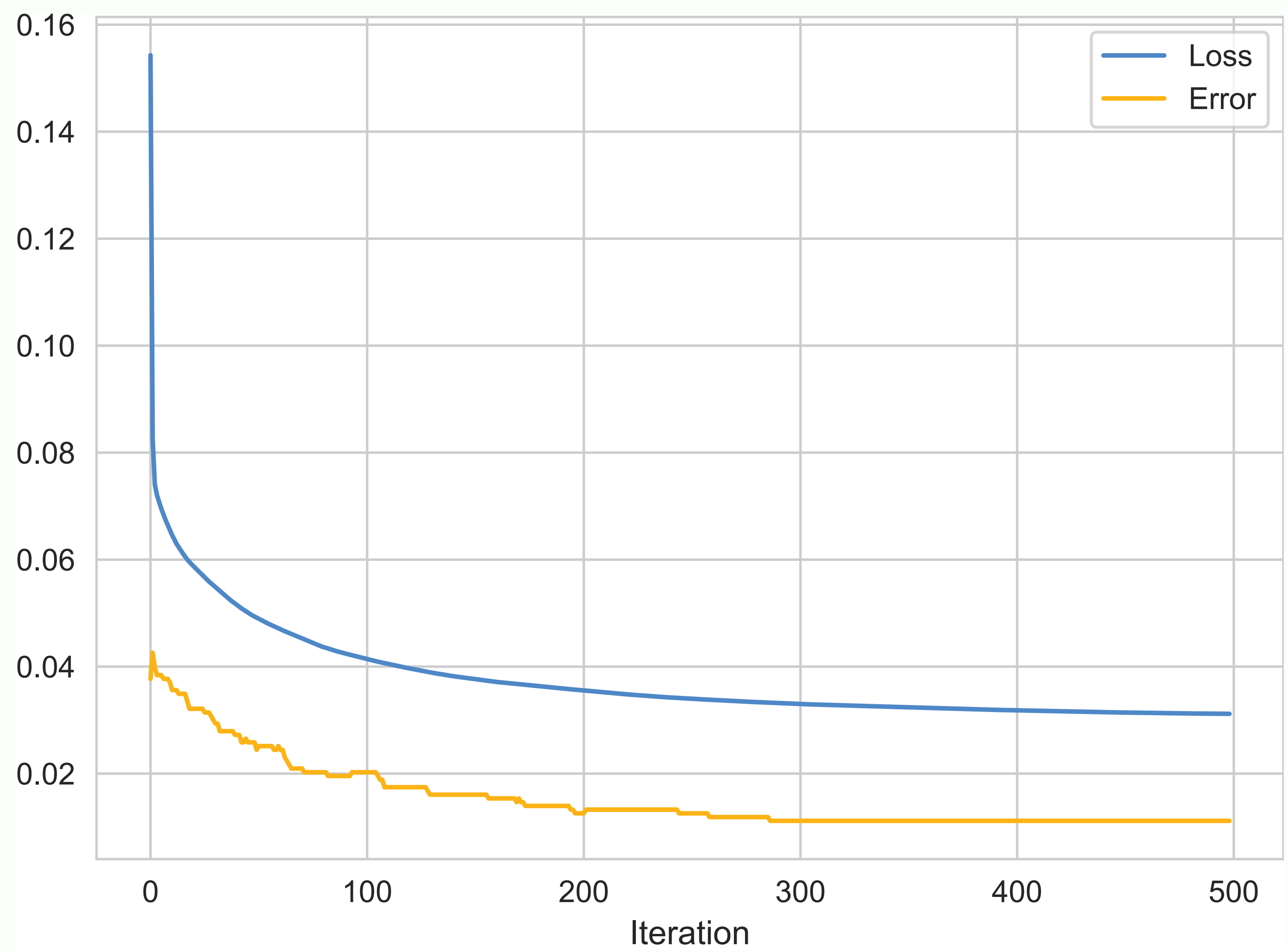
AFTER (stand.)



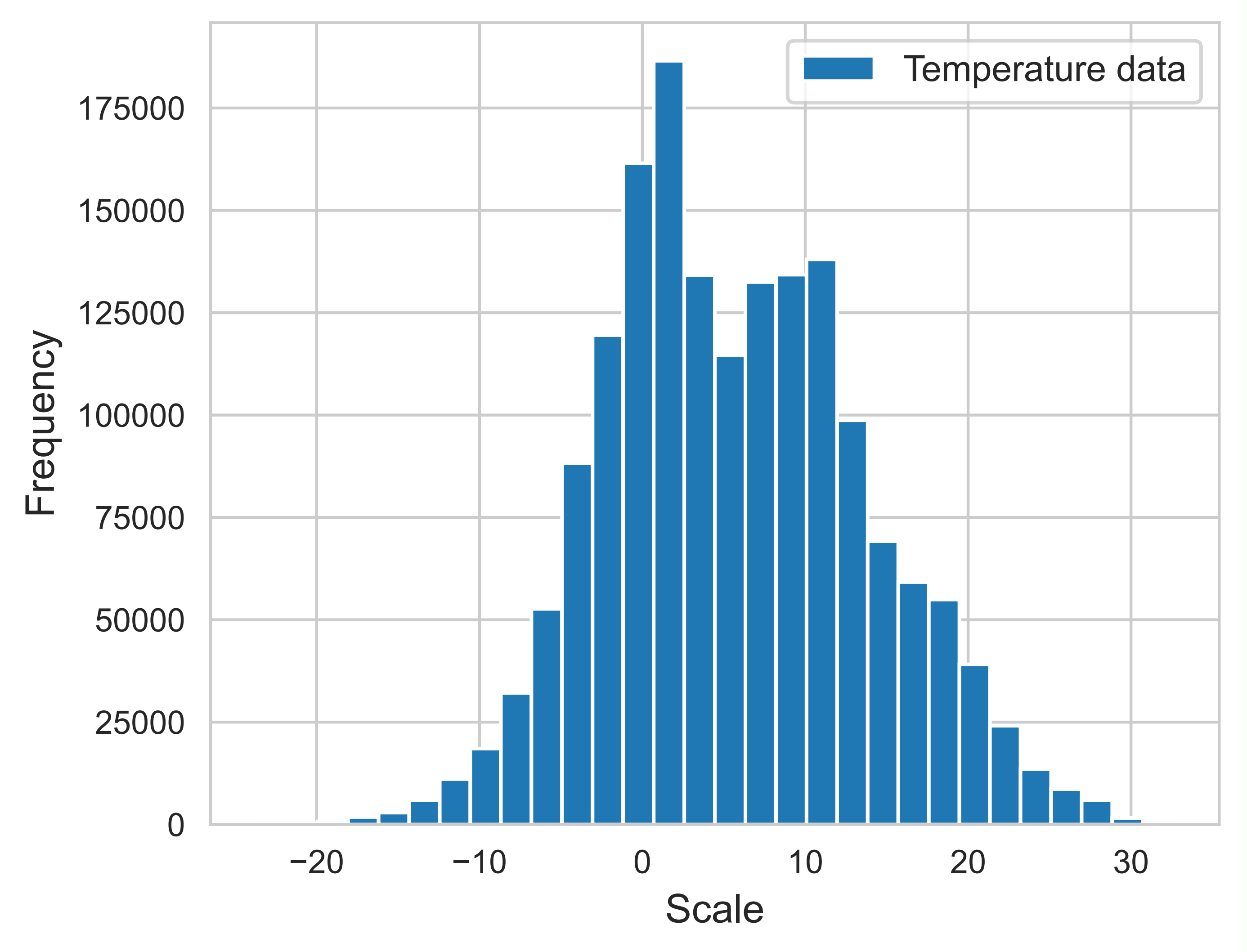
$\text{iter} = 500$

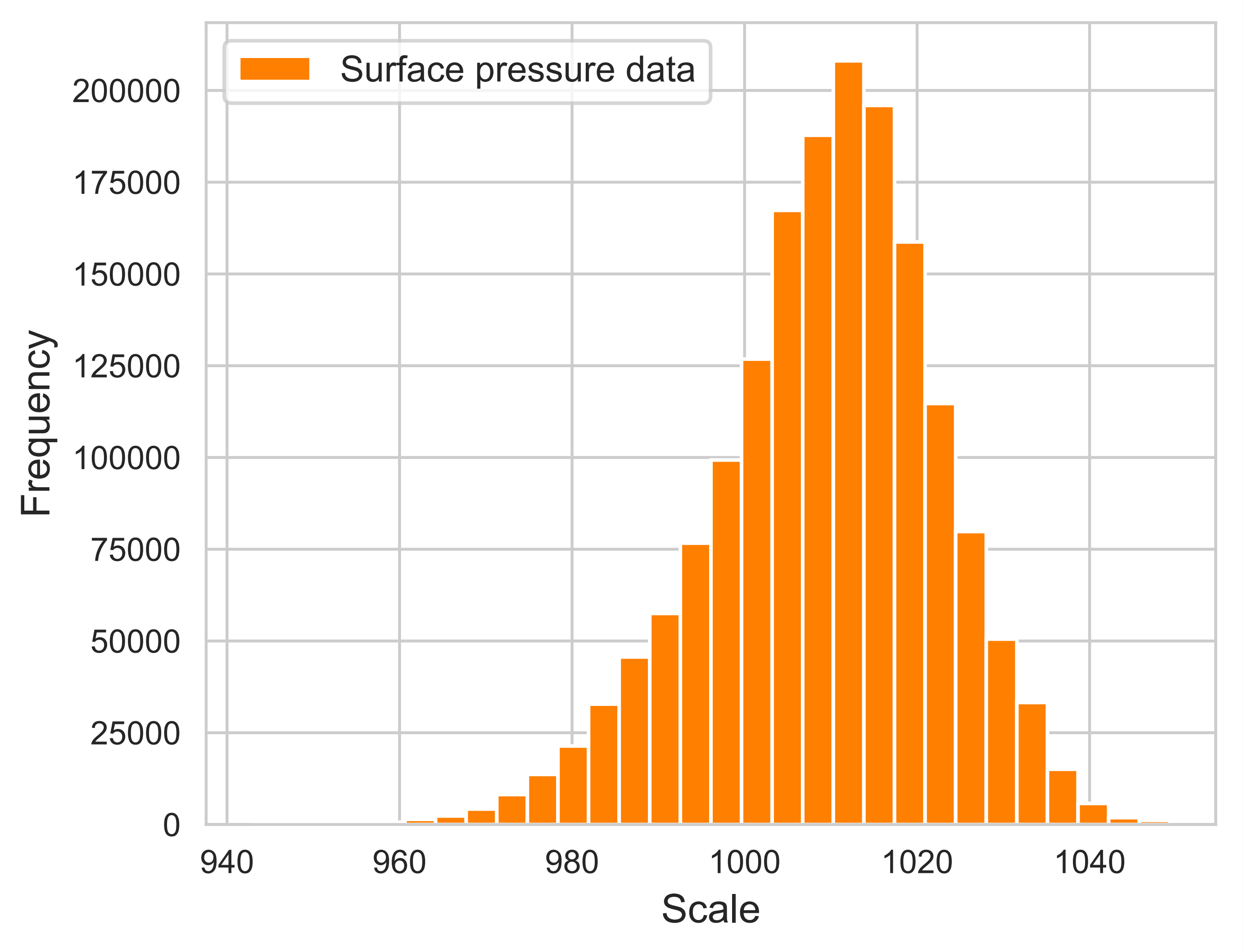
$\lambda = 10^{-4}$

$\alpha = 1$

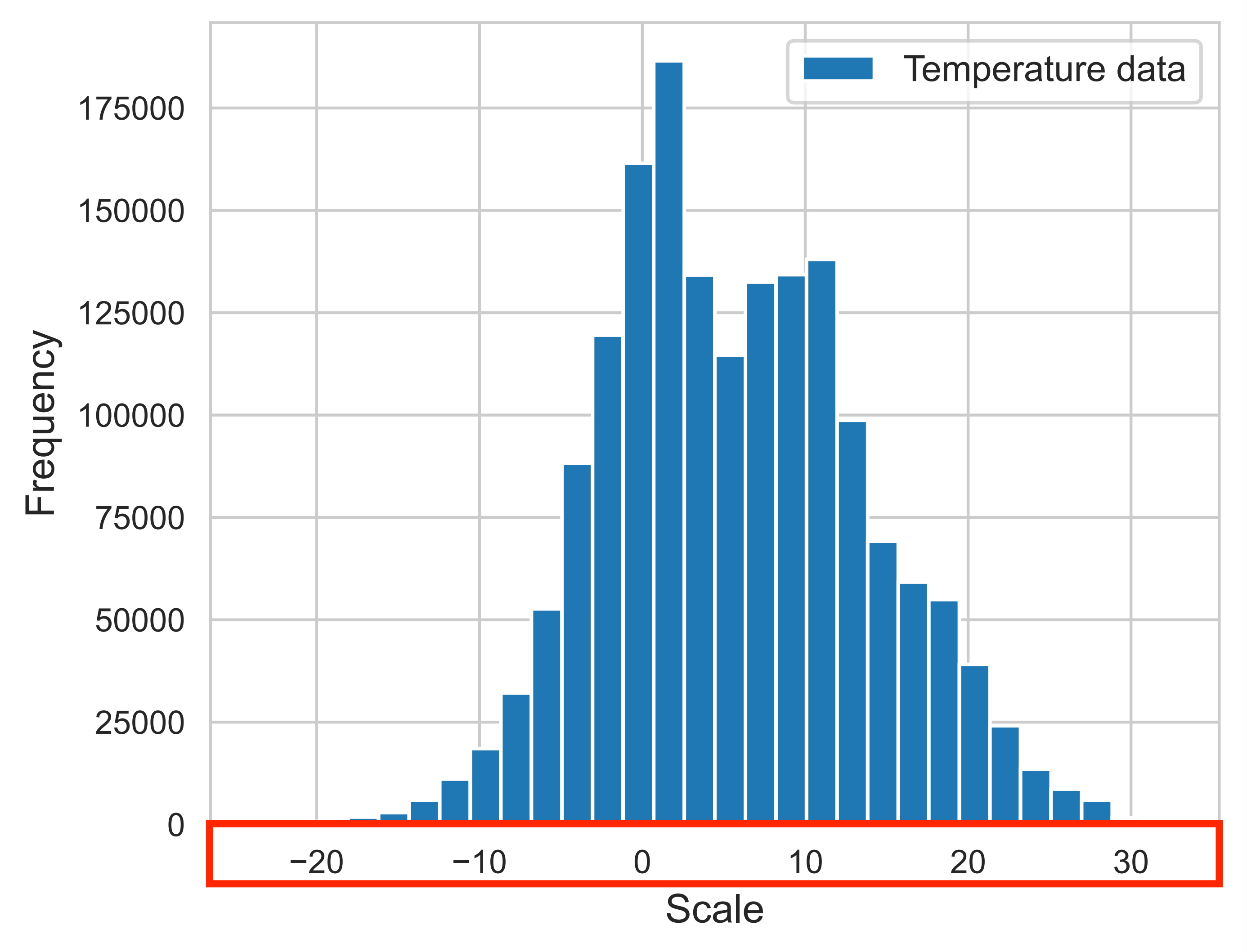


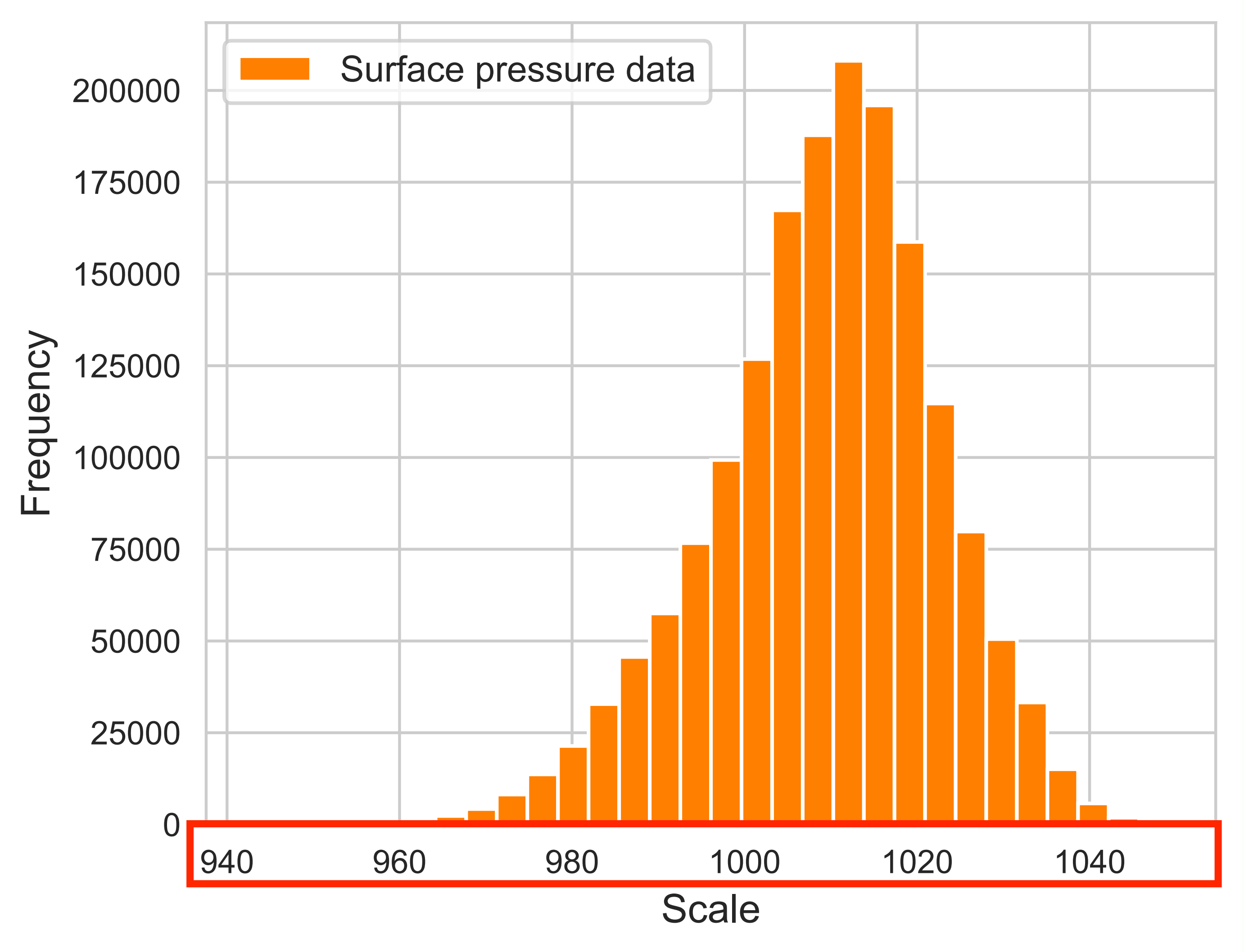
PRACTICAL OBSERVATIONS (2)



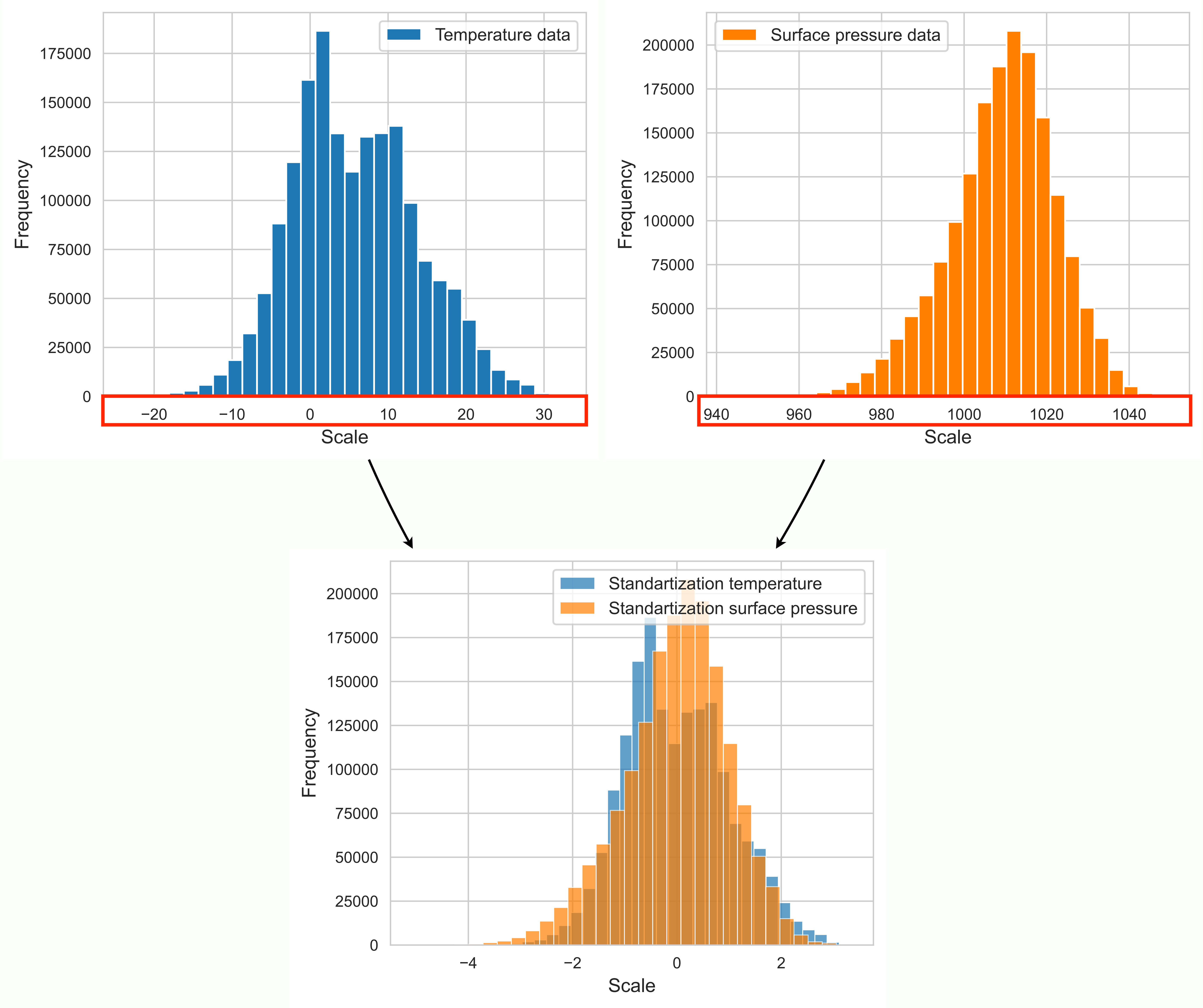


PRACTICAL OBSERVATIONS (2)

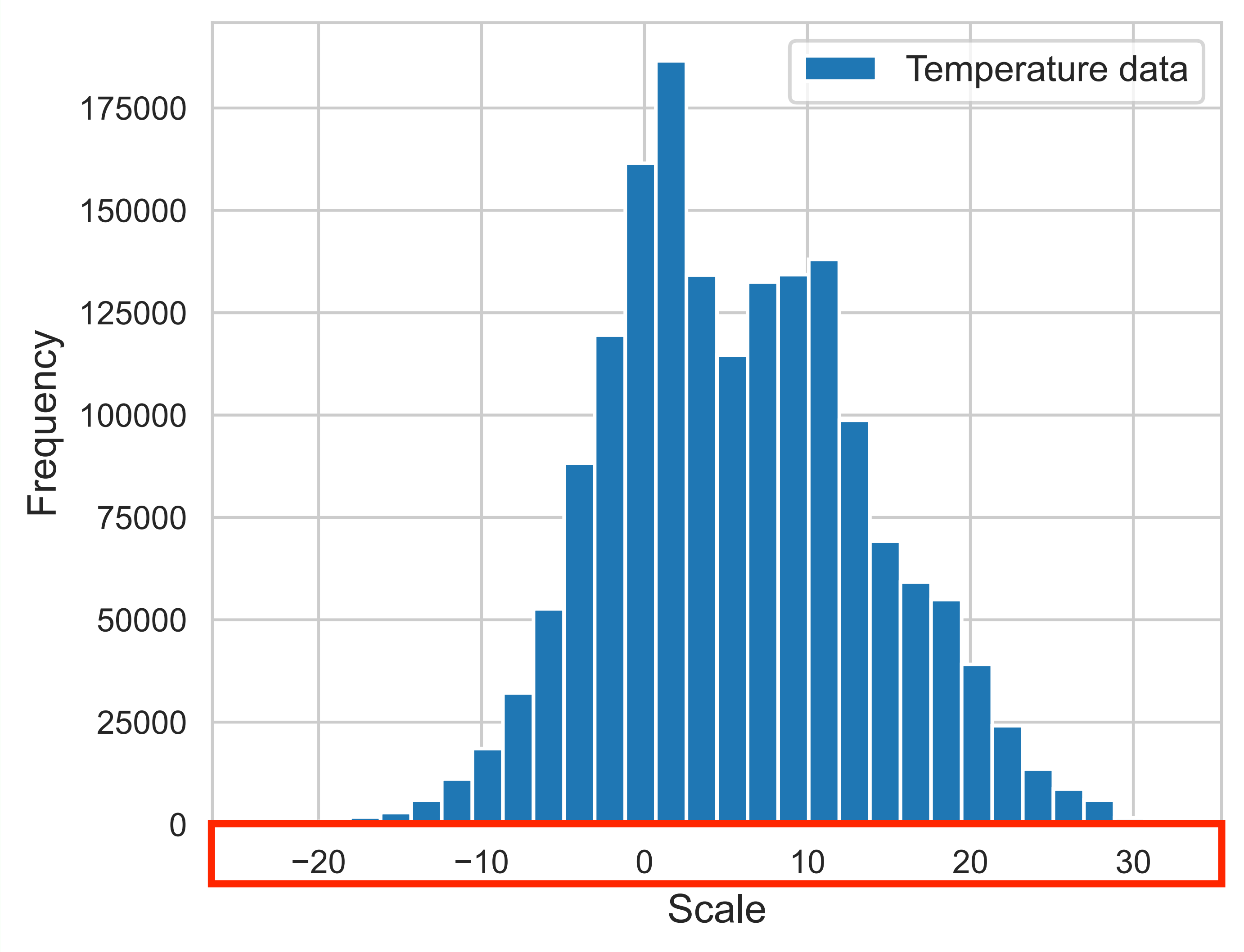


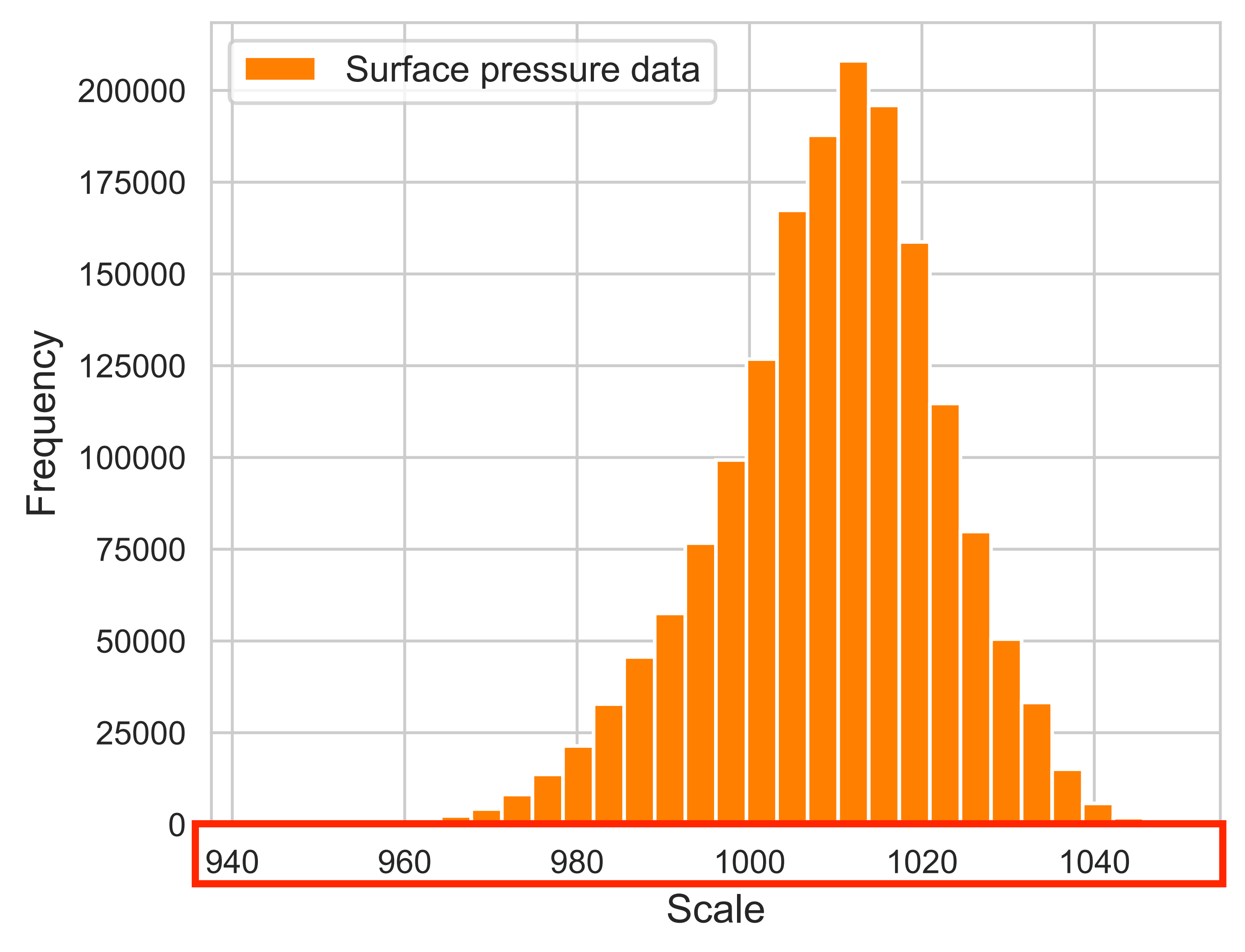


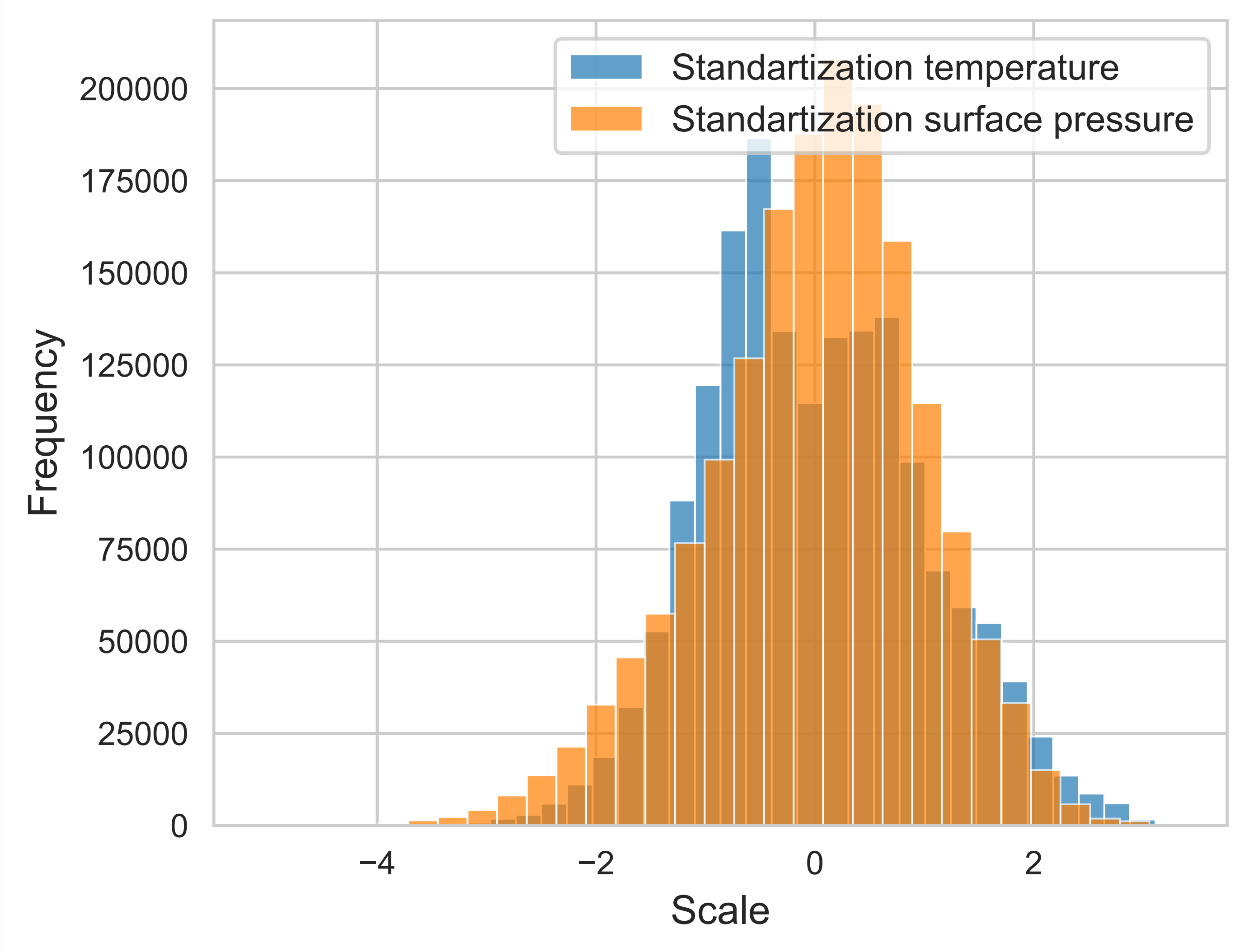
PRACTICAL OBSERVATIONS (2)



PRACTICAL OBSERVATIONS (2)

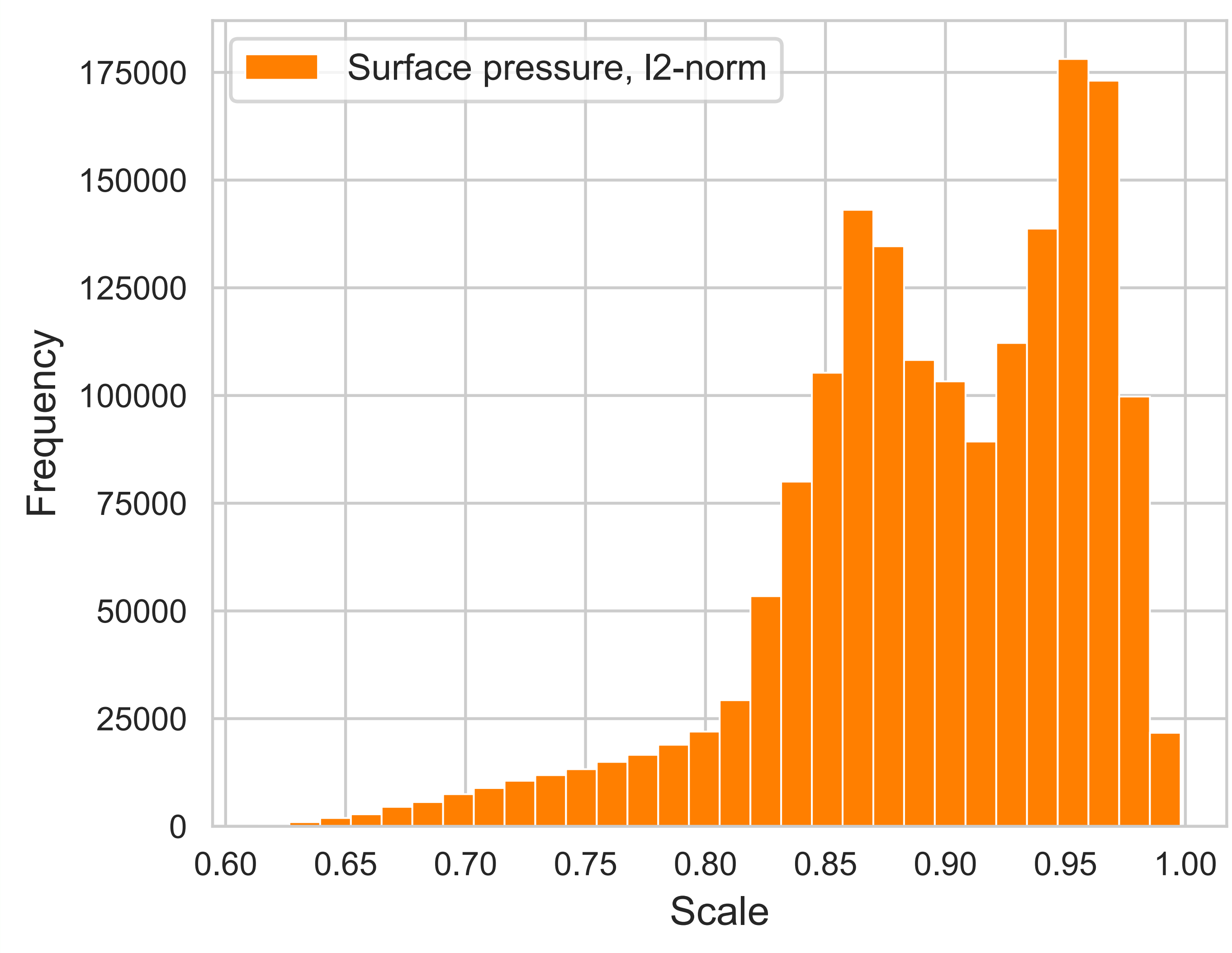






Carefully select the scaling algorithm

Rare: majority of scalers preserve the shape.



FEATURE TRANSFORMATION

Feature transformation is the process of transforming the features into a more suitable representation.

Some reasons for transformation:

- Same reasons as for feature scaling
- Reducing skewness → normal distribution shape
- Linear relationships

COMMON METHODS

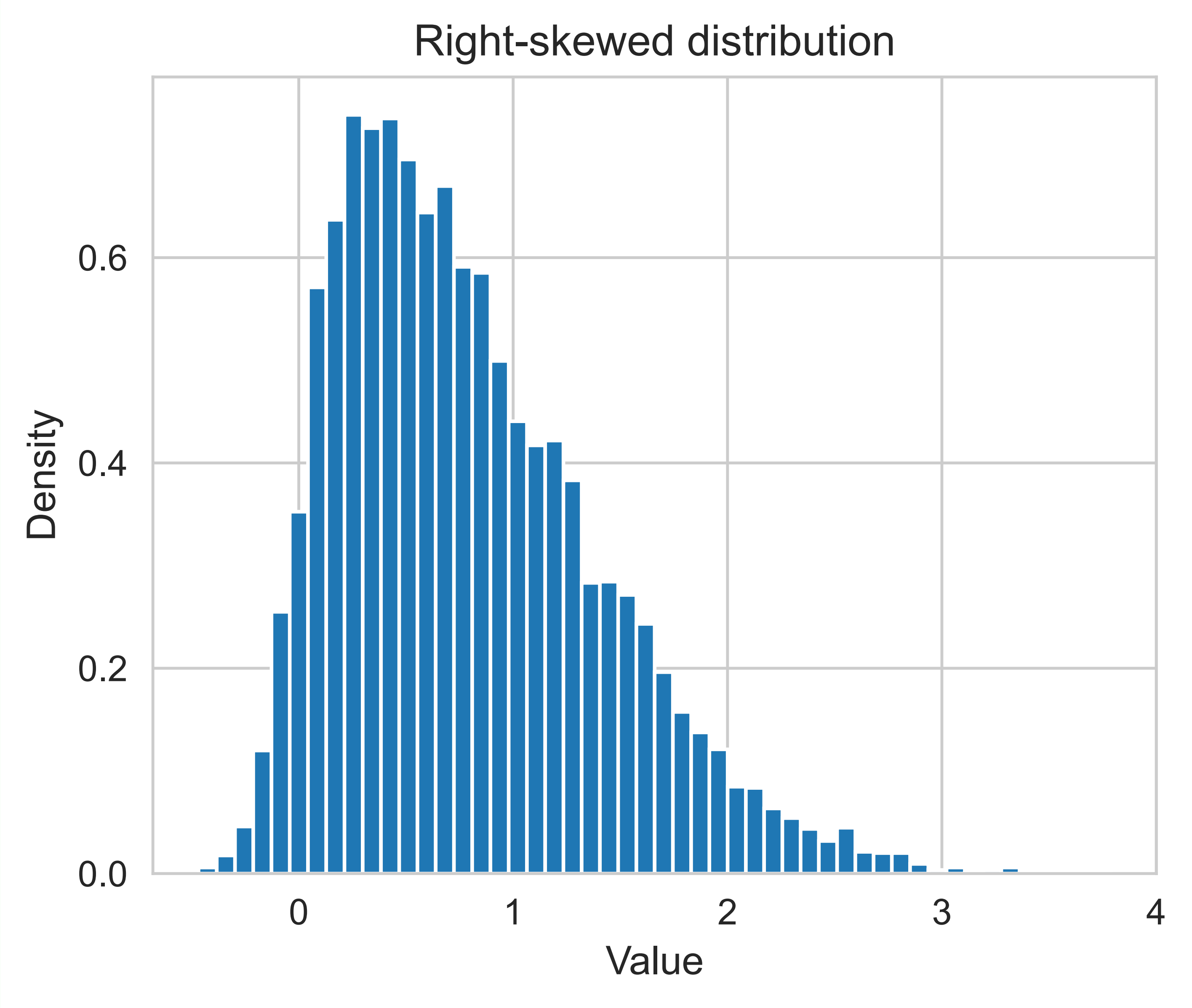
- Log
- Square/cube root
- Power (exponential, square)
- Box Cox
- Gaussian
- Encoding
- Adjustments/custom

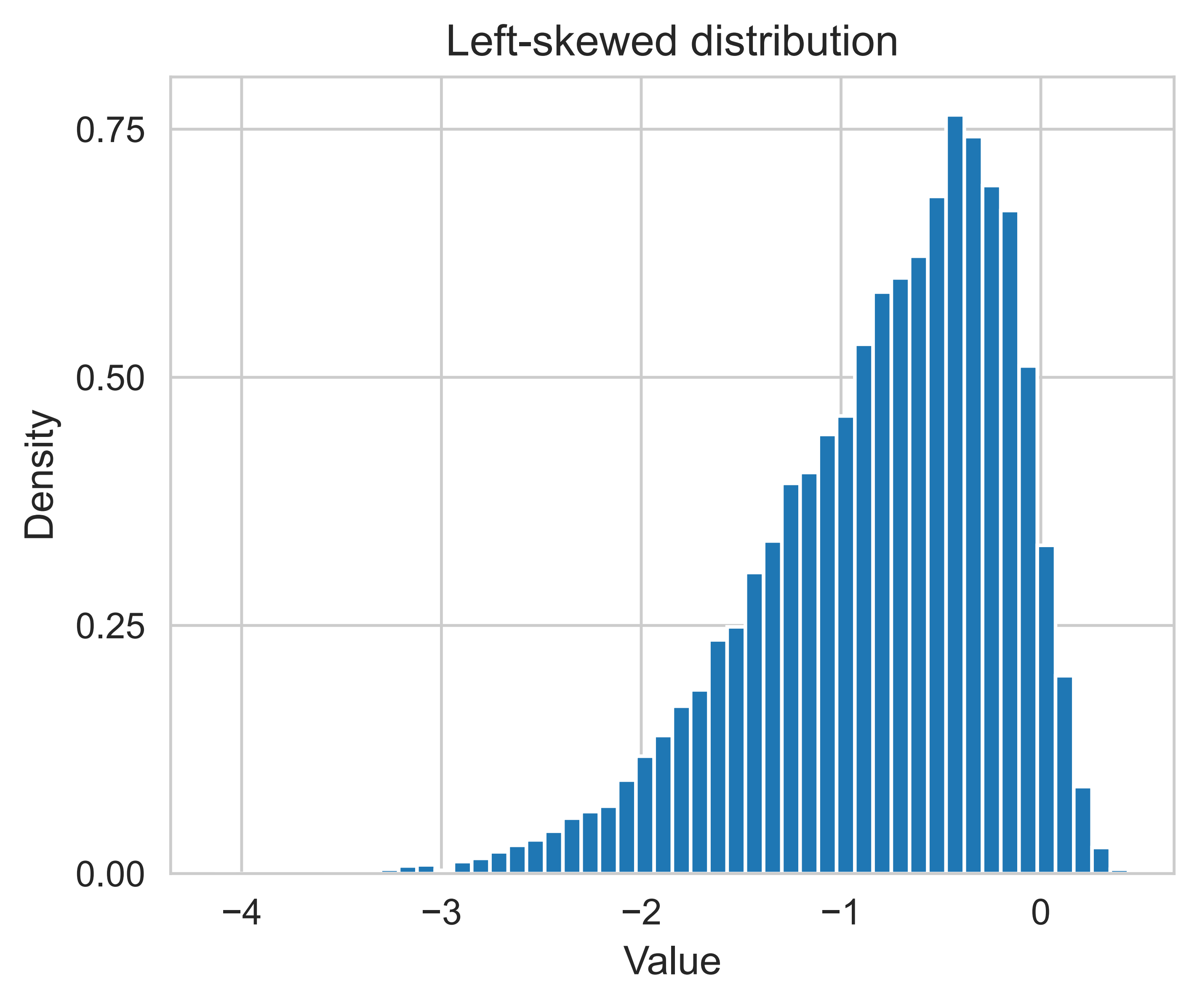
NB! Some method may change shape of the distribution.

FT: PRACTICAL TIPS

For right-skewed data (tail is on the right) common transformations include $\sqrt{x}$, $\sqrt[3]{x}$, and $\log(x)$.

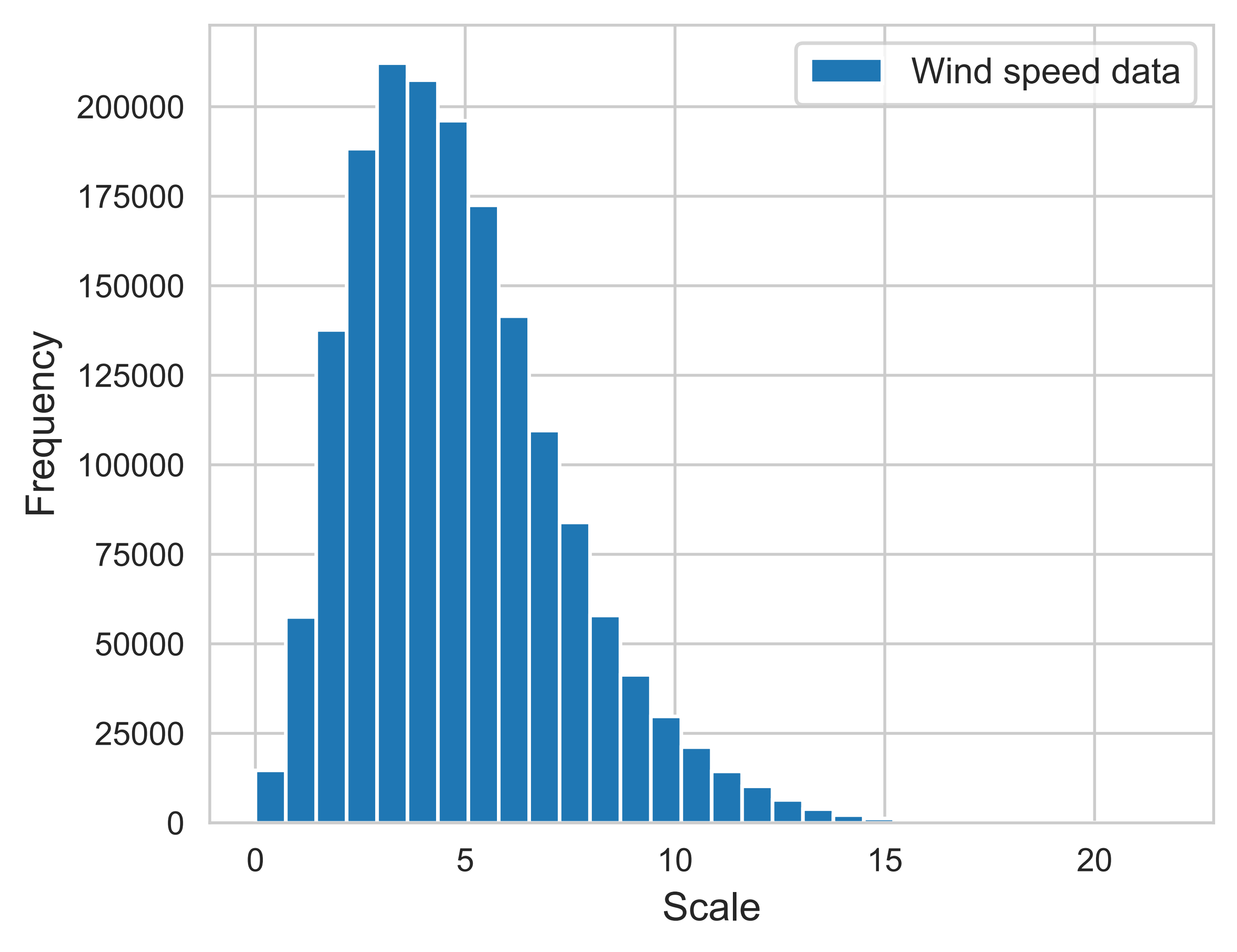
For left-skewed data (tail is on the left) common transformations include $\sqrt{\text{const} - x}$, $\sqrt[3]{\text{const} - x}$, and $\log(\text{const} - x)$.

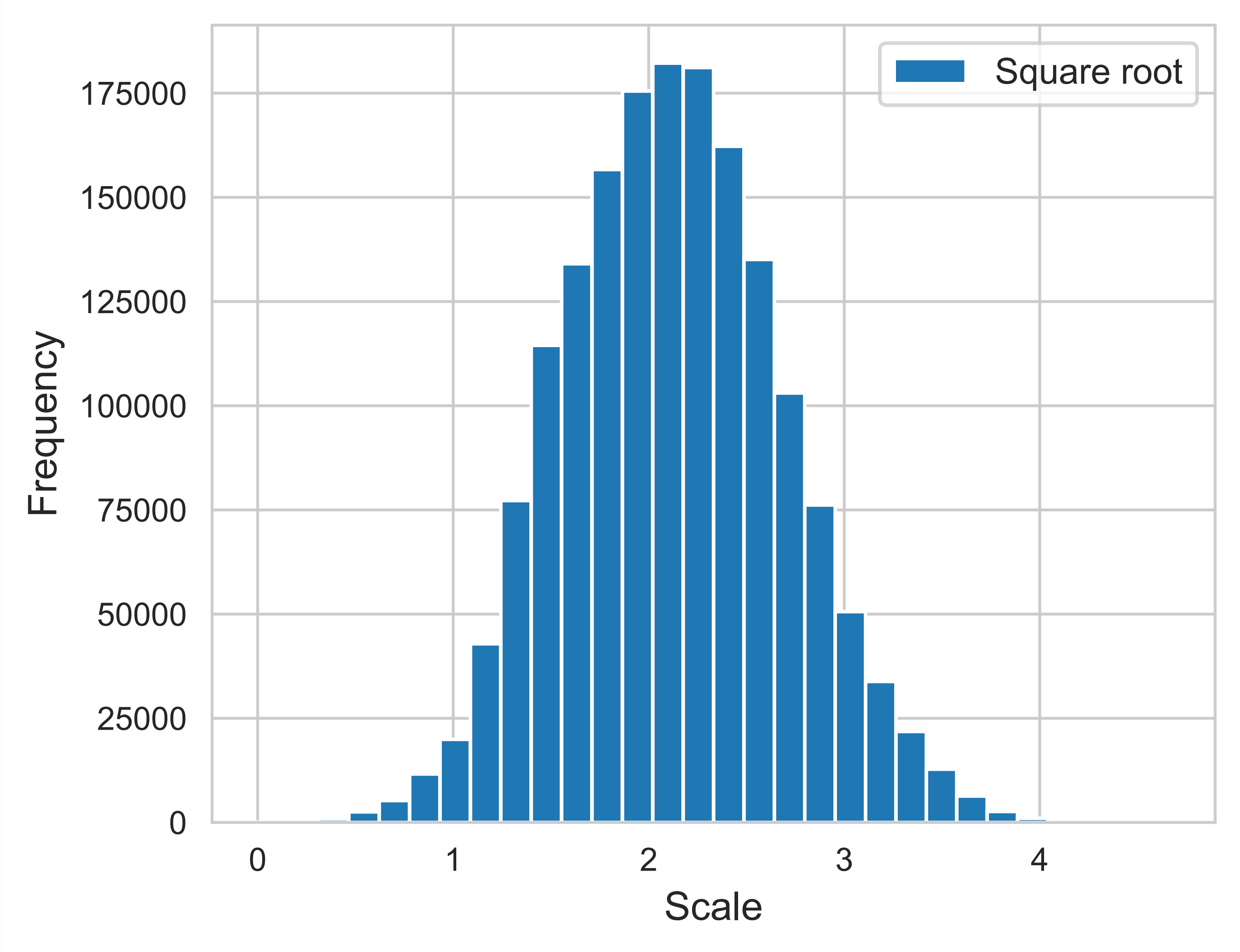


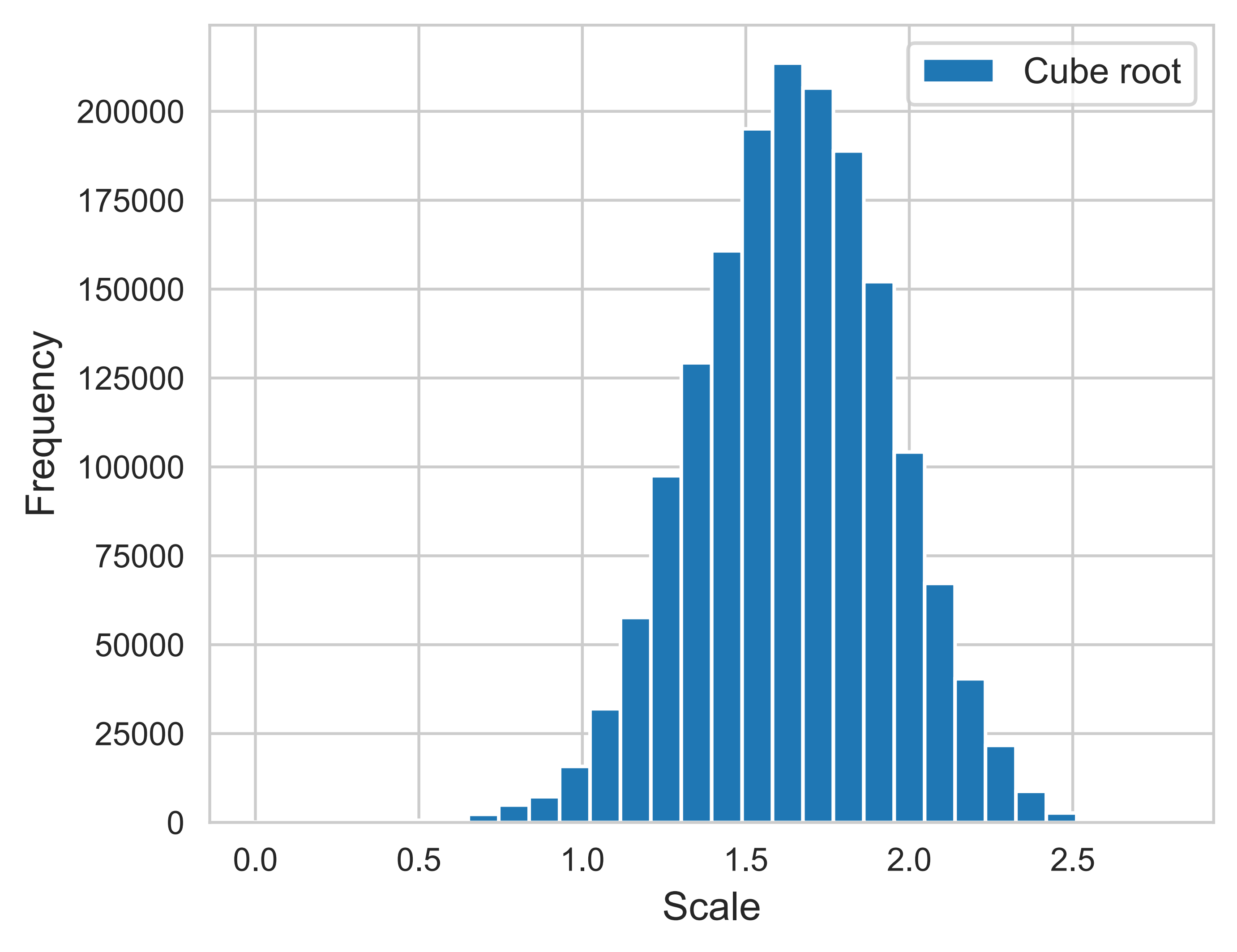


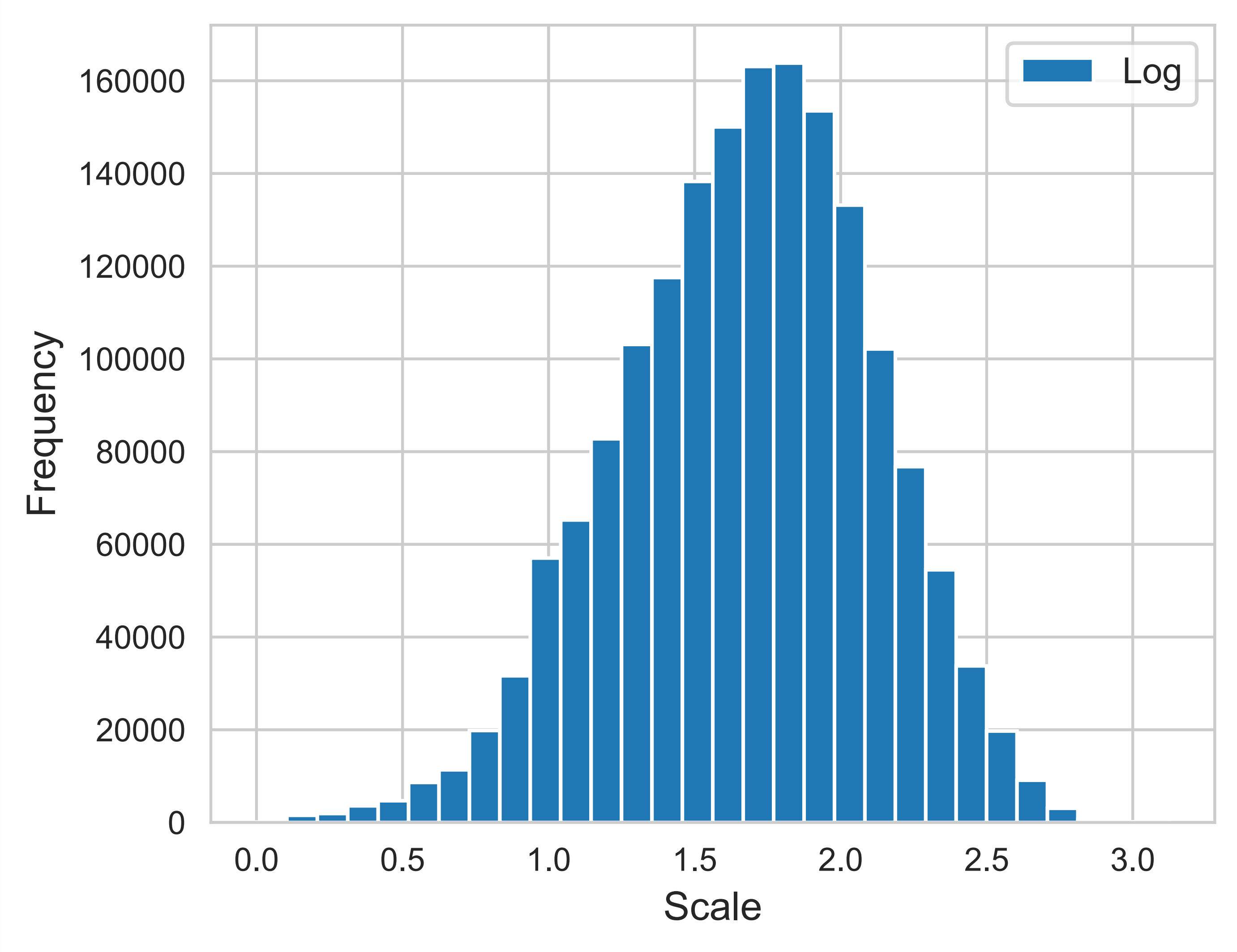
Since $\log(0)$ is undefined, when using a log transformation, a constant should be added to all values to make them all positive before transformation.

GOOD EXAMPLE

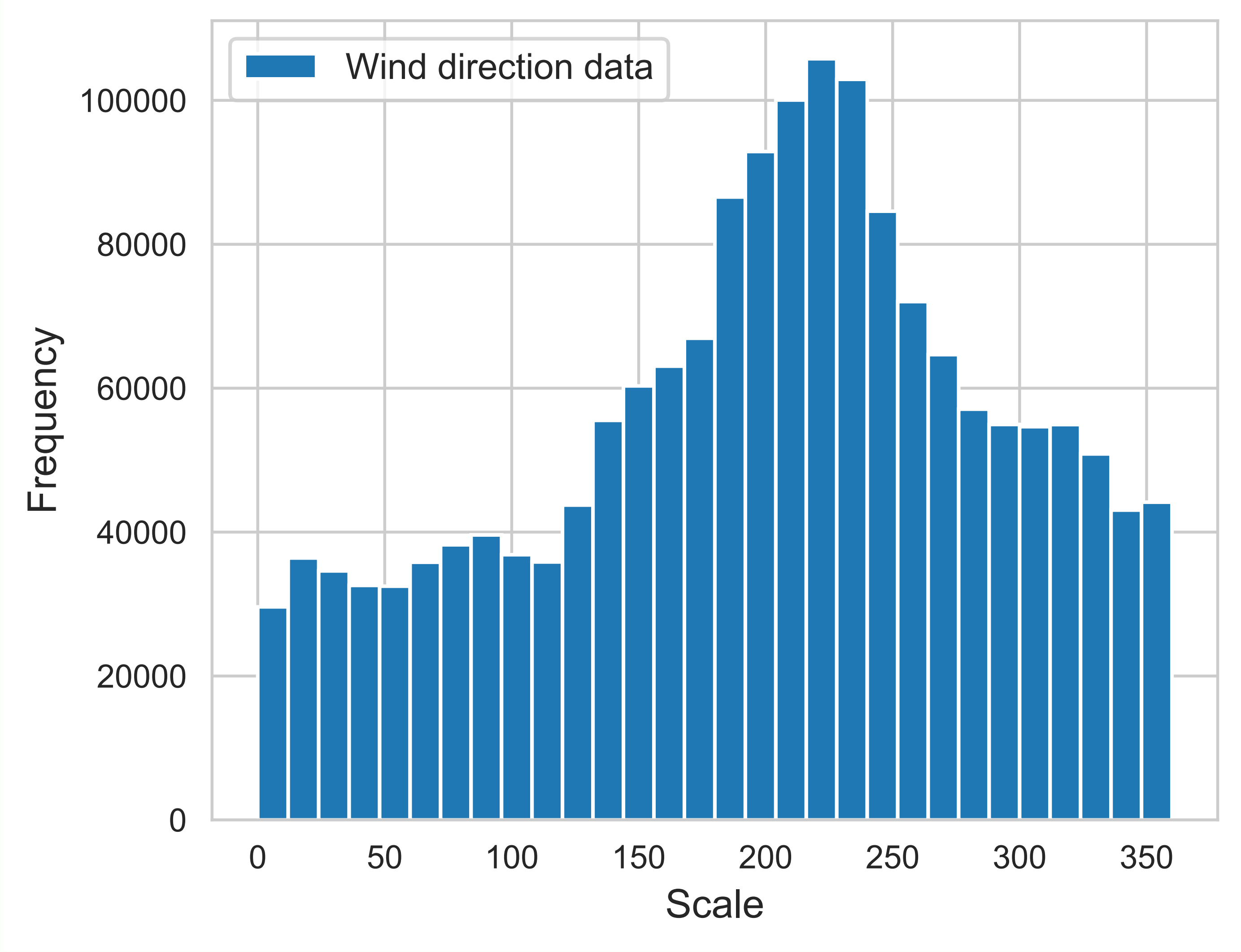




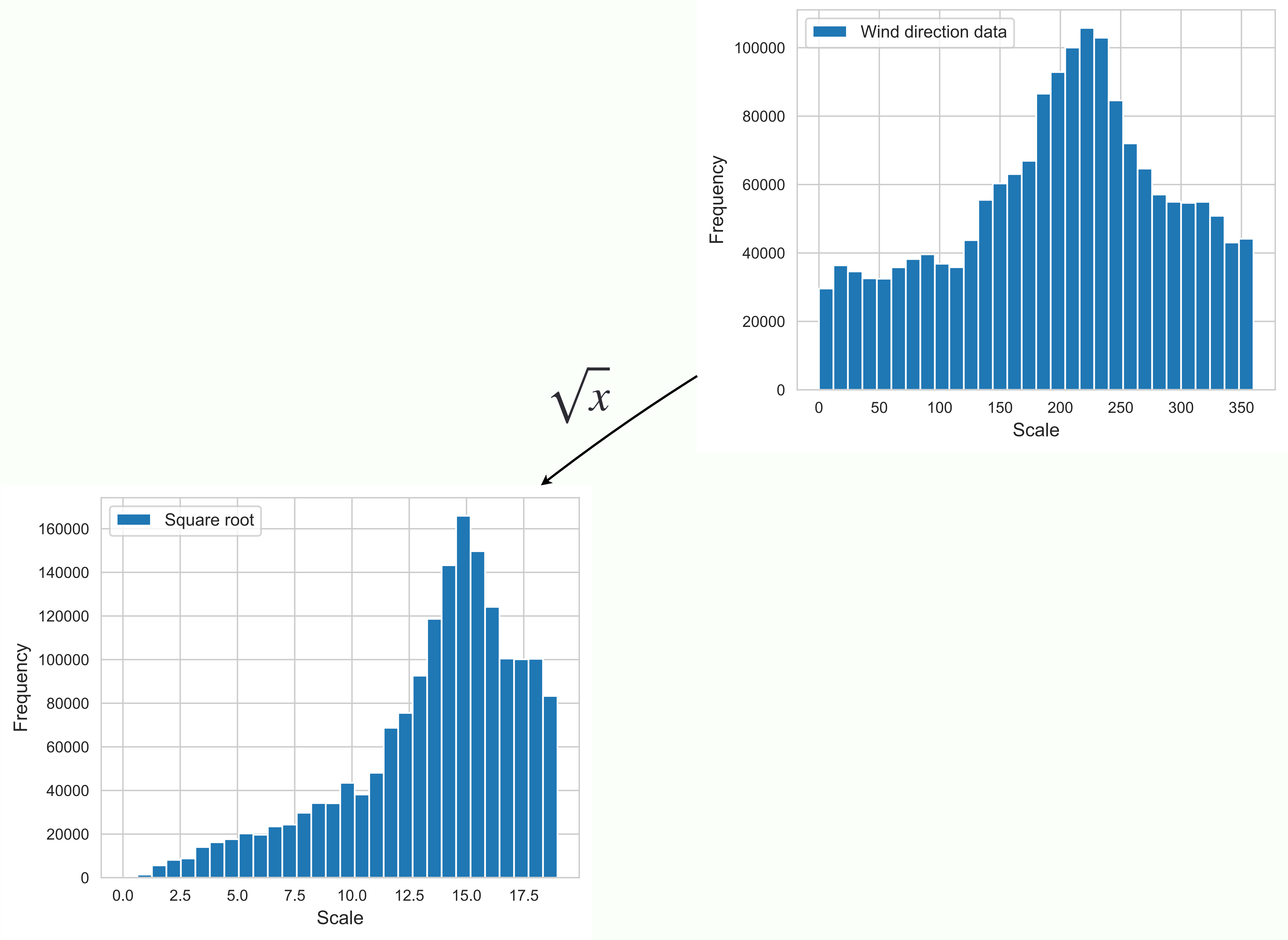




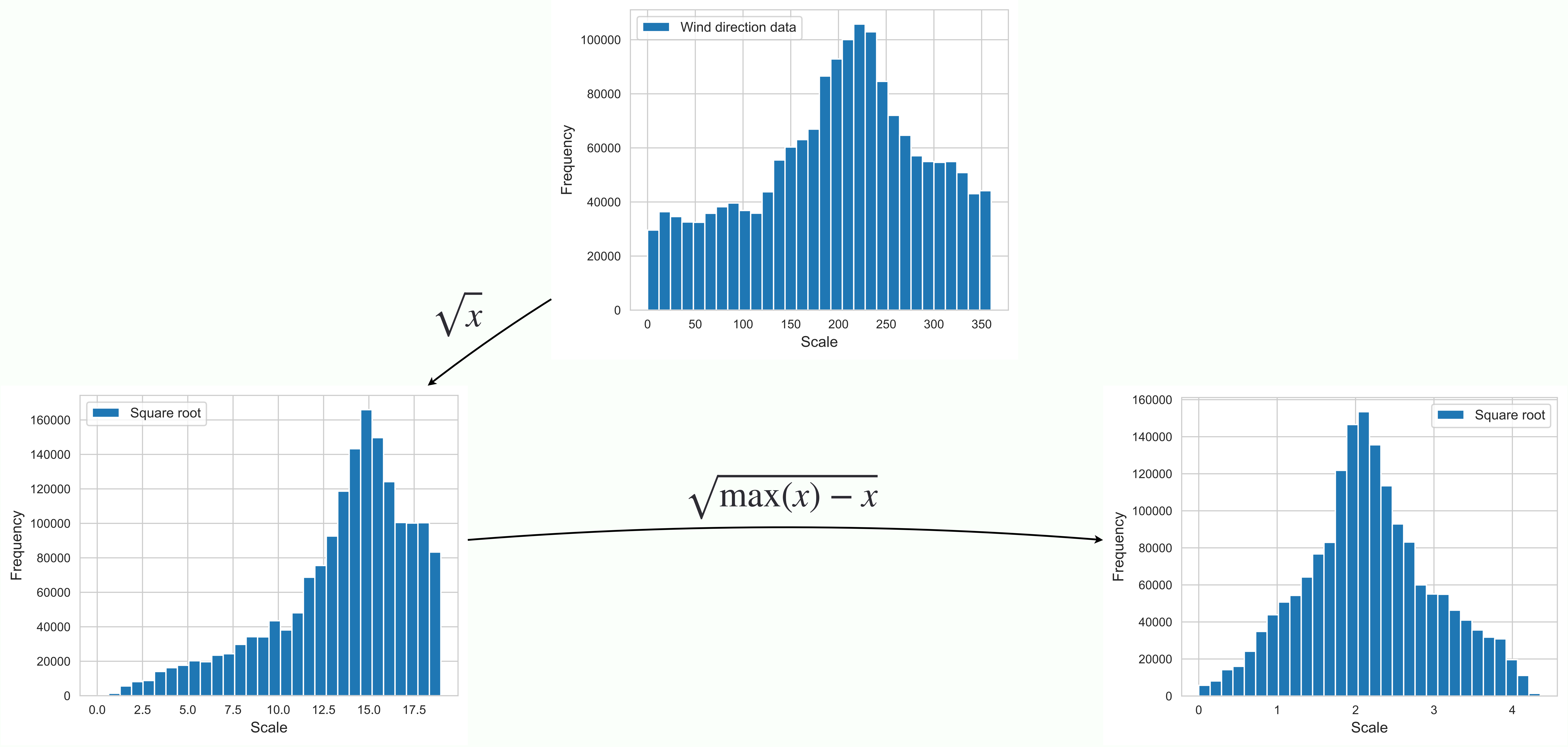
BAD EXAMPLE


Wind direction data
Frequency
100000
80000
60000
40000
20000
0
0
50
100
150
200
250
300
350
Scale

BAD EXAMPLE


Wind direction data
Frequency
Scale
$\sqrt{x}$
Square root
Frequency
Scale

BAD EXAMPLE


Wind direction data
Frequency
Scale
$\sqrt{x}$
Square root
Frequency
Scale
$\sqrt{\max(x) - x}$
Square root
Frequency
Scale

ENCODING

Transforming categorical features into a numerical representation.

Common methods:

- Numerical (frequency, mean, etc.)
- Categorical (one-hot)
- Ordinal (ordered)
- Binary
- Custom

EXAMPLE

Given a series of cloud coverage observations:

(clear, clear, half cloudy, mostly cloudy, overcast)

Observations	Numeric		Categorical (one-hot)				Ordinal	Binary
	Okta	%	Clear	Half	Mostly	Overcast		
clear	0	0	1	0	0	0	0	0
clear	0	0	1	0	0	0	0	0
half cloudy	4	[43.75,56.25)	0	1	0	0	1	1
mostly cloudy	6	[68.75,81.25)	0	0	1	0	2	1
overcast	8	100	0	0	0	1	3	1

[Accessed September 14, 2025], https://en.wikipedia.org/wiki/Okta

CUSTOM TRANSFORM

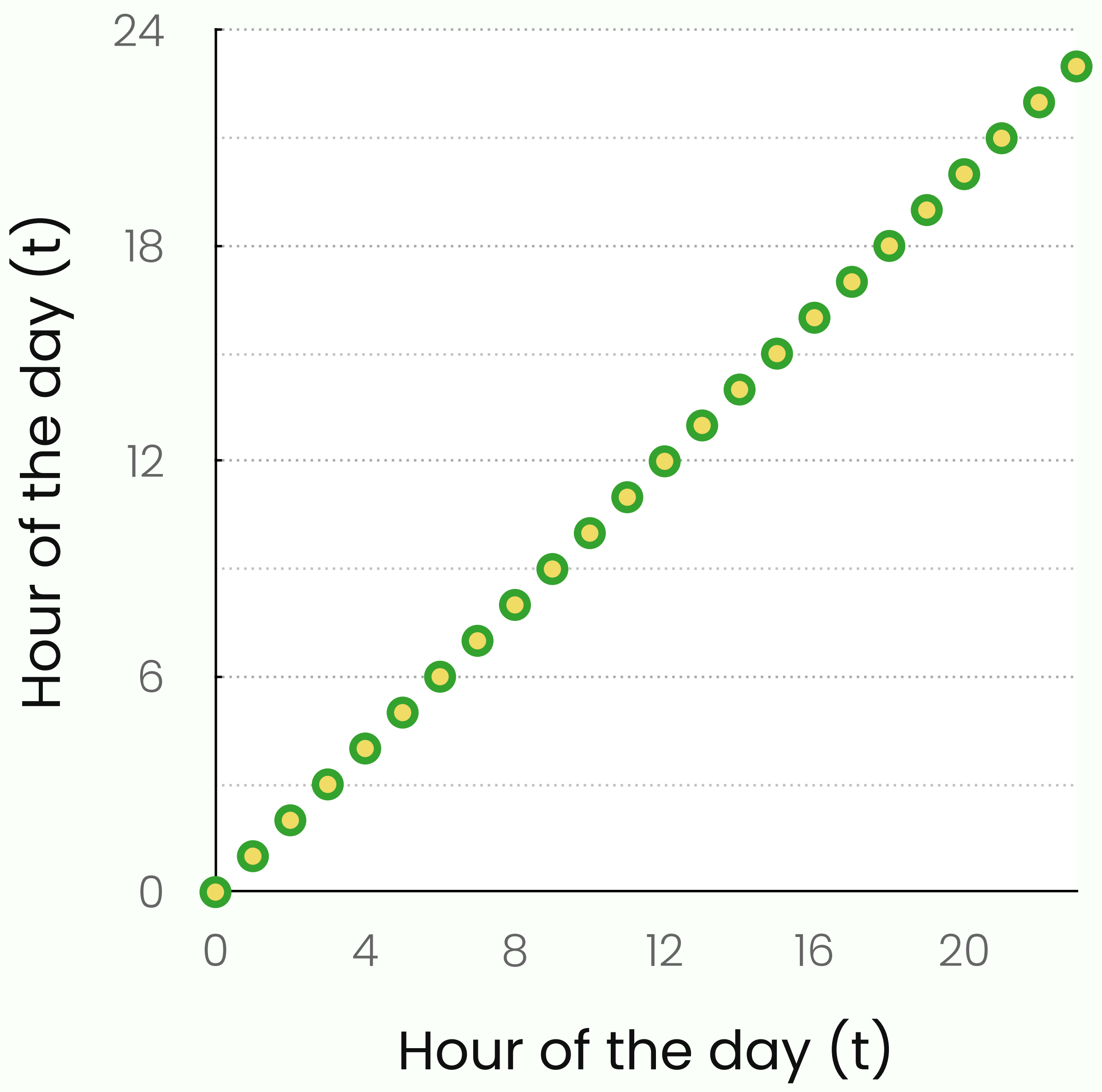
Based on the domain knowledge of the data, custom transformations can be applied to transform the data into a suitable form.

CUSTOM TRANSFORM

Based on the domain knowledge of the data, custom transformations can be applied to transform the data into a suitable form.

Hourly data with $t = 0, \ldots, 23$.

Note that the distance is always 1, but values range from 0 to 23.



CUSTOM TRANSFORM

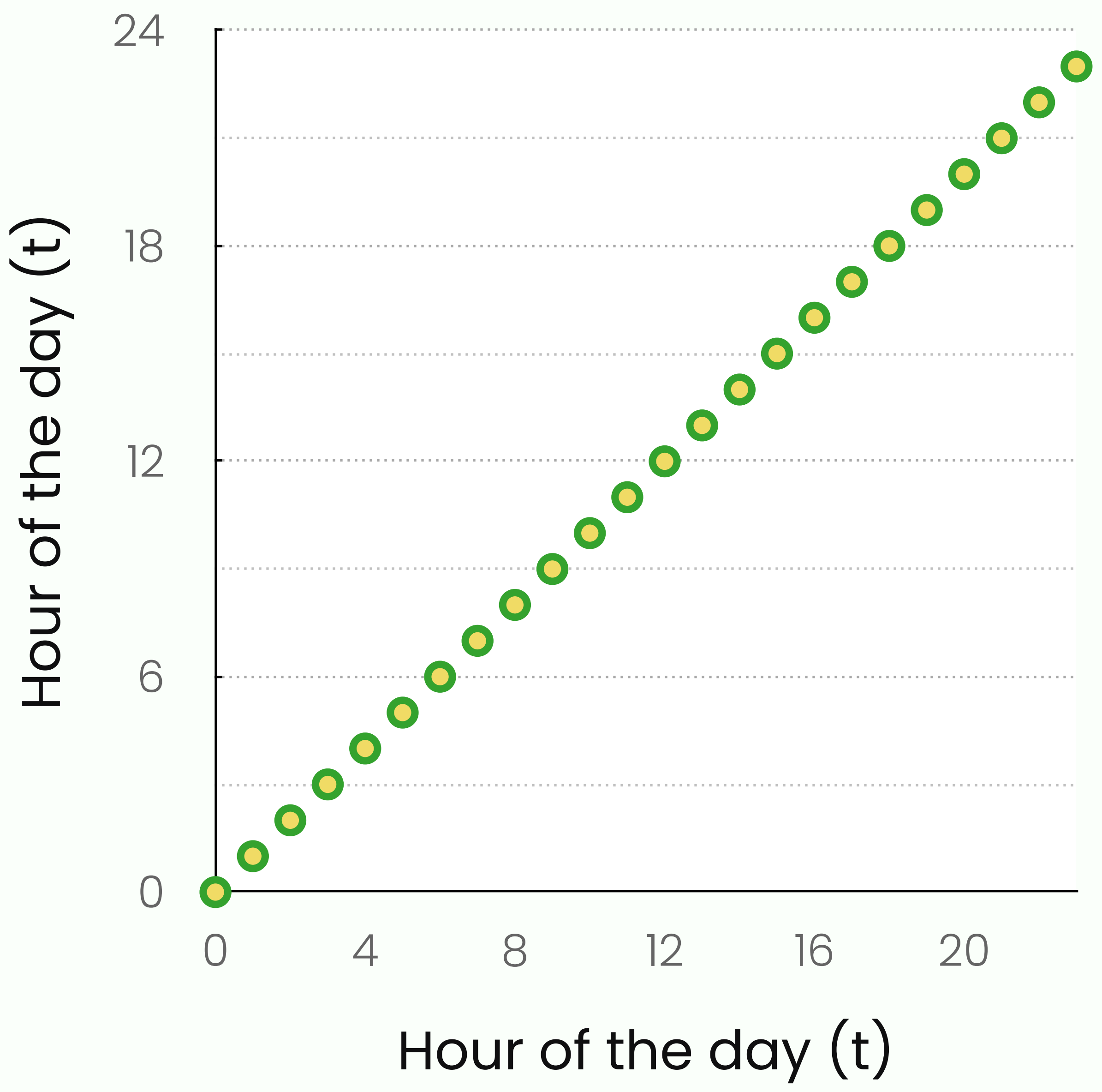
Based on the domain knowledge of the data, custom transformations can be applied to transform the data into a suitable form.

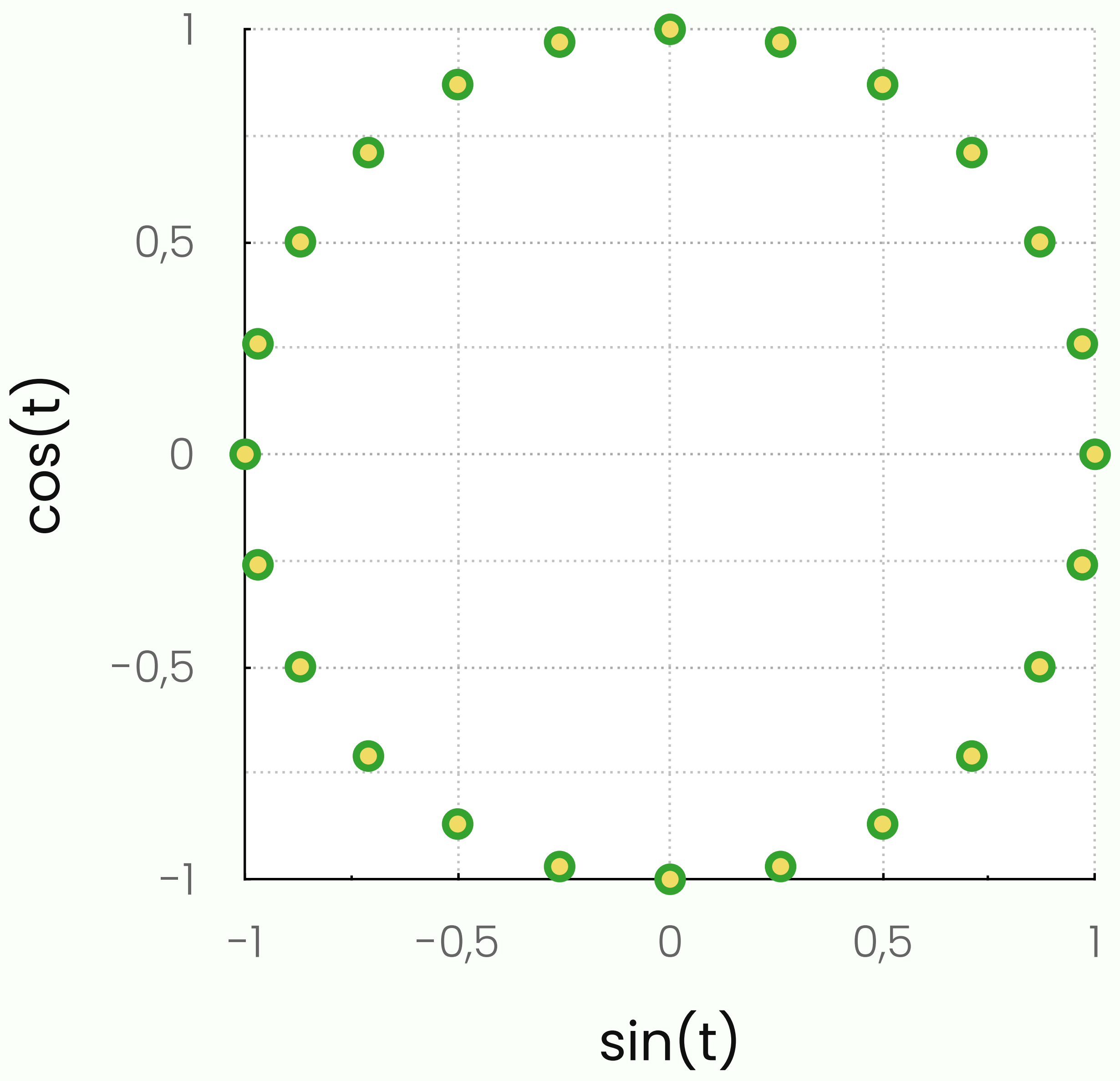
Hourly data with $t = 0,\ldots,23$.

Note that the distance is always 1, but values range from 0 to 23.

Can be transformed using $\cos(\cdot)$ and $\sin(\cdot)$ as:

$\sin(t \cdot 2\pi/24)$ and $\cos(t \cdot 2\pi/24)$.





NB! Two new features.

CALENDAR ADJUSTMENTS

If every year has exactly 365 days, what would happen to the date of the summer solstice (the longest day of the year) after 50 years?

CALENDAR ADJUSTMENTS

If every year has exactly 365 days, what would happen to the date of the summer solstice (the longest day of the year) after 50 years?

NB! The difference between the longest and shortest months is about $(31 - 28)/30 \approx 10\,\%$.

CALENDAR ADJUSTMENTS

If every year has exactly 365 days, what would happen to the date of the summer solstice (the longest day of the year) after 50 years?

NB! The difference between the longest and shortest months is about $(31 - 28)/30 \approx 10\,\%$.

For example, monthly data can be adjusted as:

$$\hat{y}_t = \frac{\text{\# of days in an average month}}{\text{\# of days in month } i} \times y_t = \frac{365.25/12}{\text{\# of days in month } i} \times y_t.$$

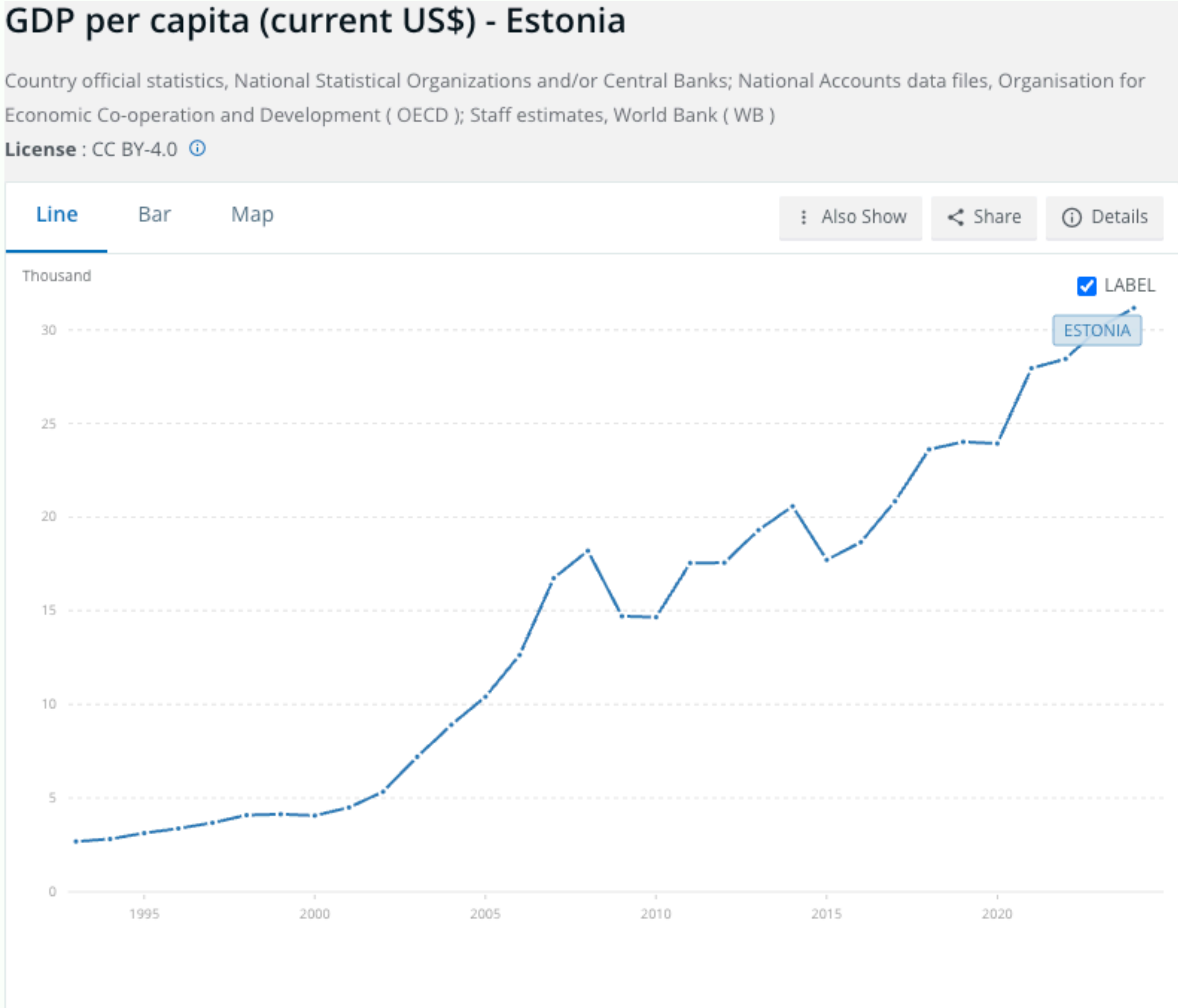
CALENDAR ADJUSTMENTS: EXAMPLES

Financial quarters are not always equal in length (Q4 has holidays and an extra day in leap year).

Months vary between 28-31 days → raw averages (e.g., *visits per month*) can be misleading unless normalised to days.

Heating demand (season) depends on real winter, not a specific day in the calendar.

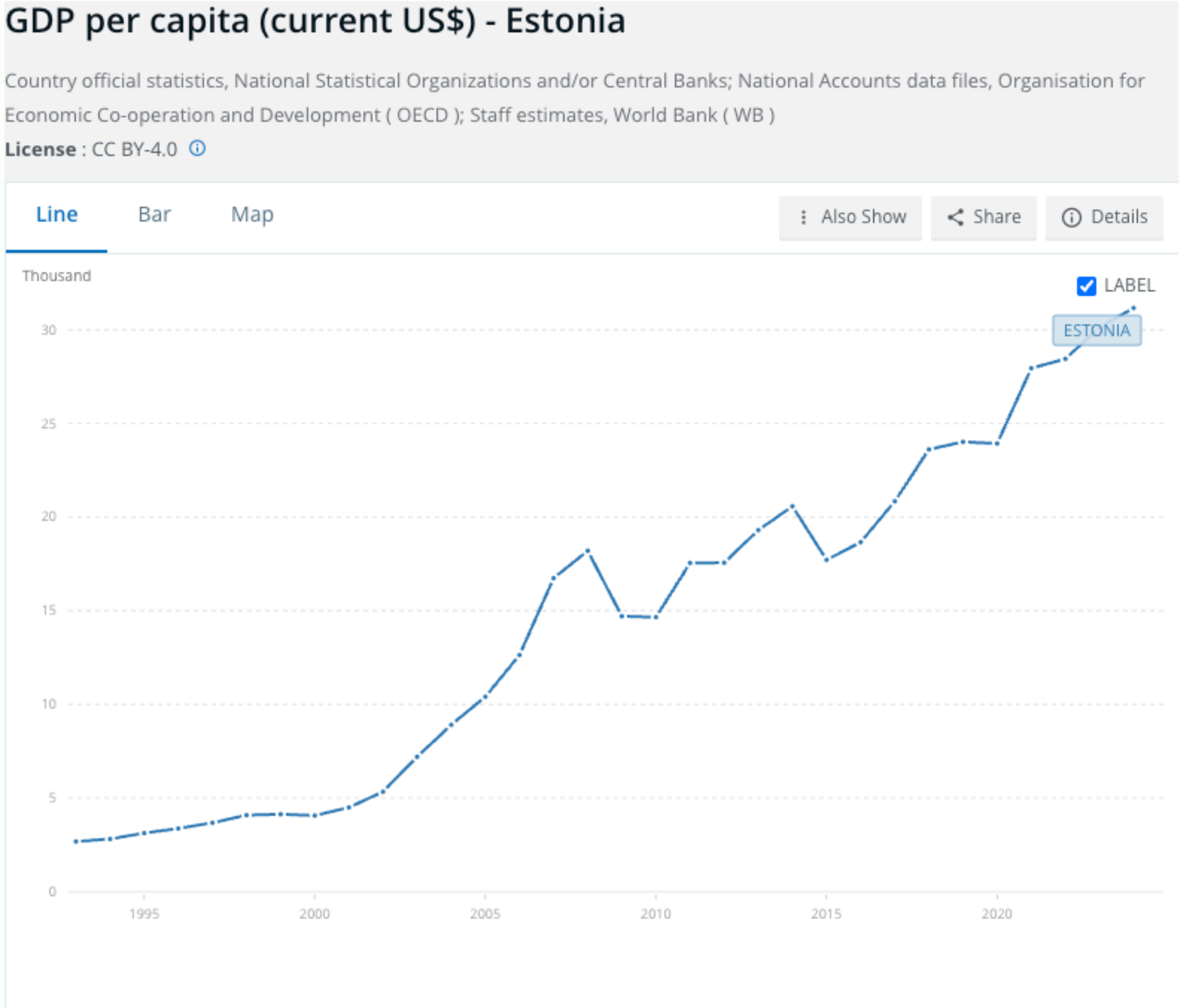
OTHER ADJUSTMENTS



Year	GDP Nominal (Current USD)	GDP Real (Inflation adj.)	GDP Change	GDP per capita	Pop. Change	Population
2023	$41,291,245,222	$27,574,843,698	−3.02%	$20,169	1.27%	1,367,196
2022	$38,376,046,175	$28,434,426,605	0.06%	$21,061	1.38%	1,350,091
2021	$37,204,563,051	$28,417,254,674	7.15%	$21,338	0.16%	1,331,749
2020	$31,820,771,494	$26,519,902,199	−2.88%	$19,945	0.21%	1,329,669
2019	$31,873,748,770	$27,307,395,616	3.73%	$20,581	0.37%	1,326,822
2018	$31,222,632,741	$26,326,305,514	3.7%	$19,915	0.34%	1,321,965
2017	$27,469,461,919	$25,387,061,800	5.63%	$19,270	0.12%	1,317,425
2016	$24,561,027,788	$24,032,886,893	3.09%	$18,264	0.1%	1,315,849
2015	$23,311,847,751	$23,311,847,751	1.84%	$17,733	0.01%	1,314,576
2014	$27,055,689,003	$22,891,073,863	3.32%	$17,415	−0.26%	1,314,452
2013	$25,451,032,781	$22,154,913,170	1.76%	$16,811	−0.36%	1,317,919
2012	$23,237,406,116	$21,772,480,719	3.67%	$16,462	−0.36%	1,322,616
2011	$23,303,915,795	$21,001,308,237	7.61%	$15,822	−0.31%	1,327,358
2010	$19,524,355,419	$19,516,679,024	2.45%	$14,658	−0.23%	1,331,448
2009	$19,633,984,440	$19,050,002,369	−14.63%	$14,275	−0.19%	1,334,528
2008	$24,342,935,404	$22,314,690,793	−5.13%	$16,689	−0.27%	1,337,074

[Accessed September 14], https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2024&locations=EE&start=1993&view=chart
https://www.worldometers.info/gdp/estonia-gdp/

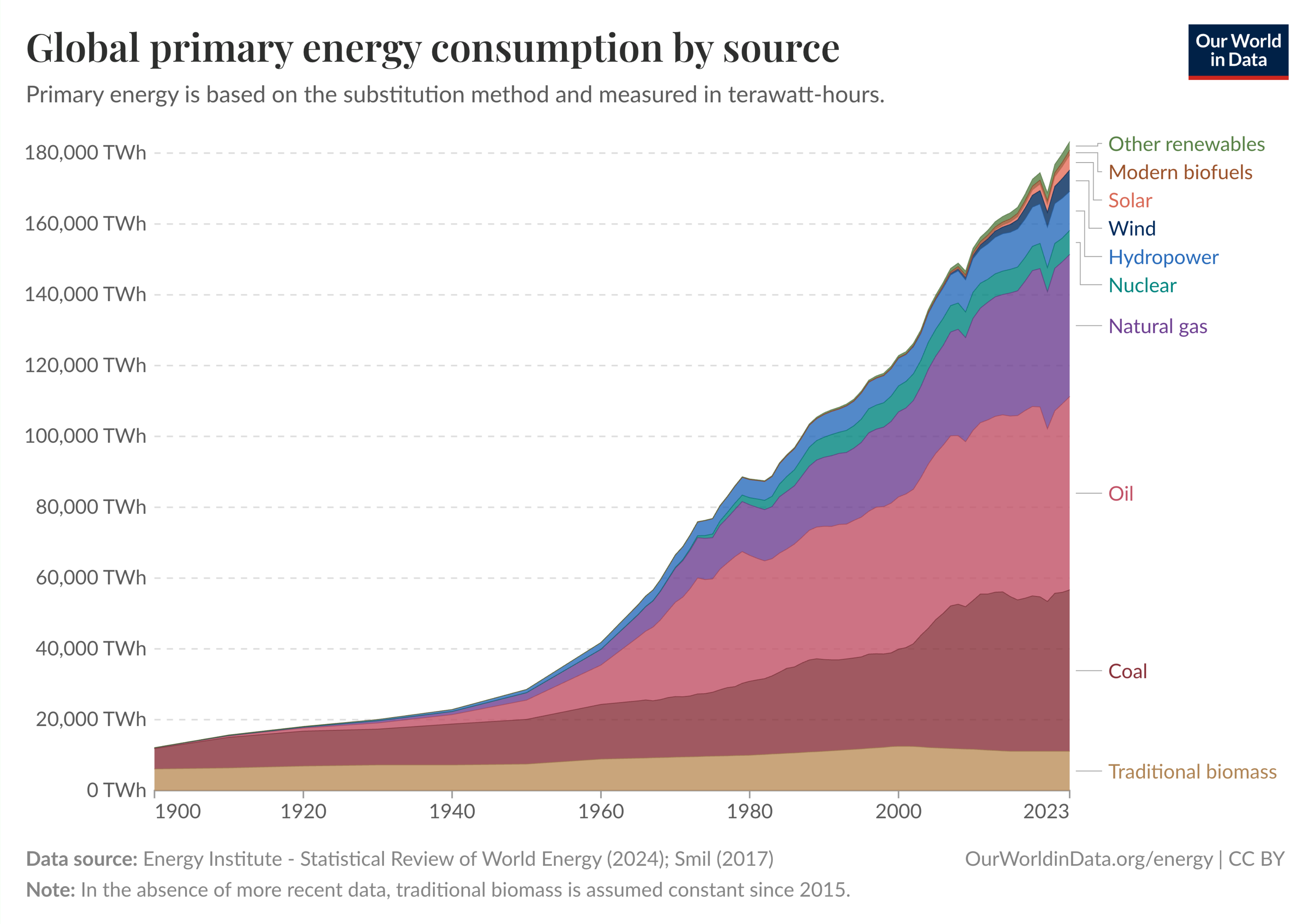
OTHER ADJUSTMENTS



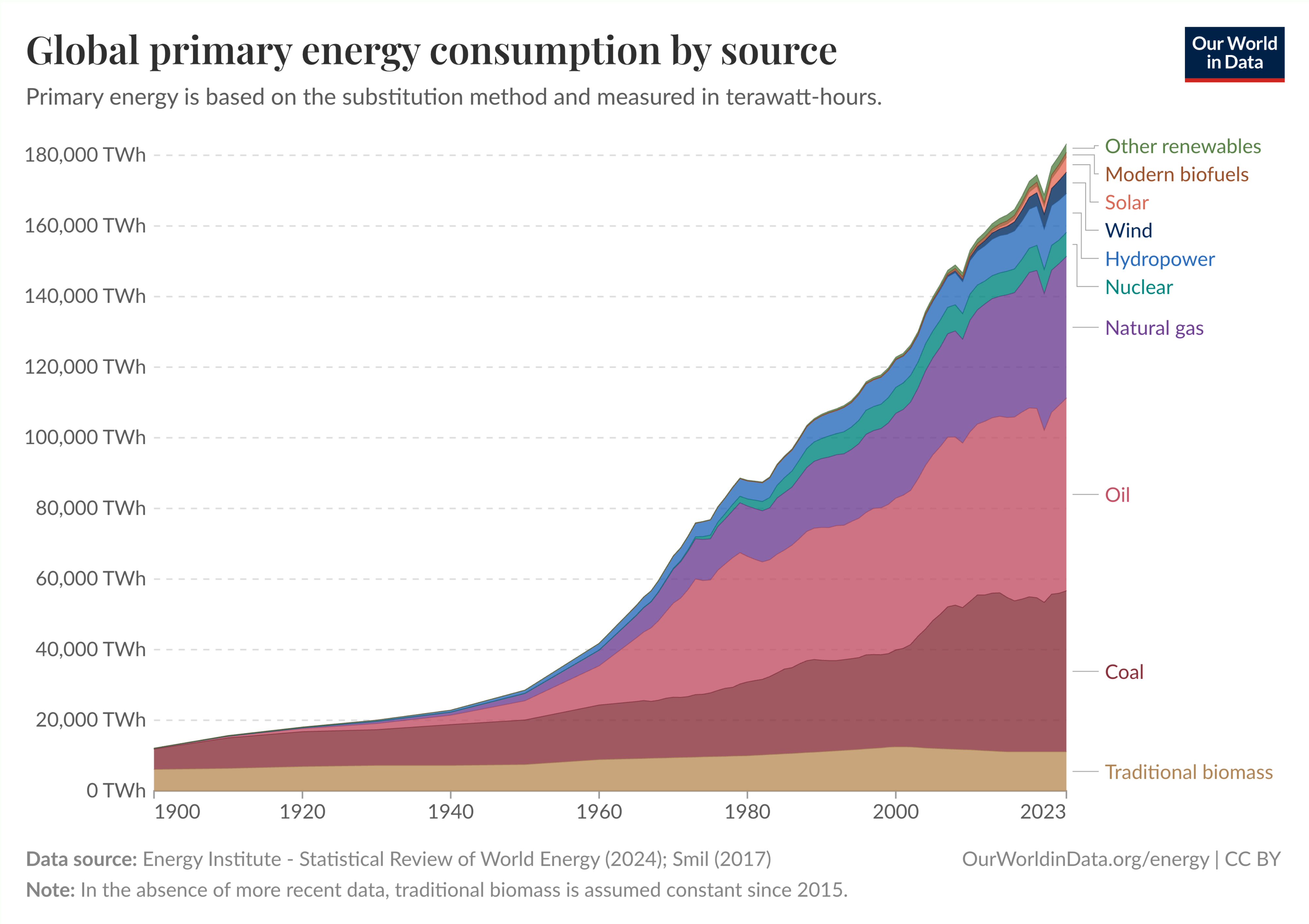
Year	GDP Nominal (Current USD)	GDP Real (Inflation adj.)	GDP Change	GDP per capita	Pop. Change	Population
2023	$41,291,245,222	$27,574,843,698	−3.02%	$20,169	1.27%	1,367,196
2022	$38,376,046,175	$28,434,426,605	0.06%	$21,061	1.38%	1,350,091
2021	$37,204,563,051	$28,417,254,674	7.15%	$21,338	0.16%	1,331,749
2020	$31,820,771,494	$26,519,902,199	−2.88%	$19,945	0.21%	1,329,669
2019	$31,873,748,770	$27,307,395,616	3.73%	$20,581	0.37%	1,326,822
2018	$31,222,632,741	$26,326,305,514	3.7%	$19,915	0.34%	1,321,965
2017	$27,469,461,919	$25,387,061,800	5.63%	$19,270	0.12%	1,317,425
2016	$24,561,027,788	$24,032,886,893	3.09%	$18,264	0.1%	1,315,849
2015	$23,311,847,751	$23,311,847,751	1.84%	$17,733	0.01%	1,314,576
2014	$27,055,689,003	$22,891,073,863	3.32%	$17,415	−0.26%	1,314,452
2013	$25,451,032,781	$22,154,913,170	1.76%	$16,811	−0.36%	1,317,919
2012	$23,237,406,116	$21,772,480,719	3.67%	$16,462	−0.36%	1,322,616
2011	$23,303,915,795	$21,001,308,237	7.61%	$15,822	−0.31%	1,327,358
2010	$19,524,355,419	$19,516,679,024	2.45%	$14,658	−0.23%	1,331,448
2009	$19,633,984,440	$19,050,002,369	−14.63%	$14,275	−0.19%	1,334,528
2008	$24,342,935,404	$22,314,690,793	−5.13%	$16,689	−0.27%	1,337,074

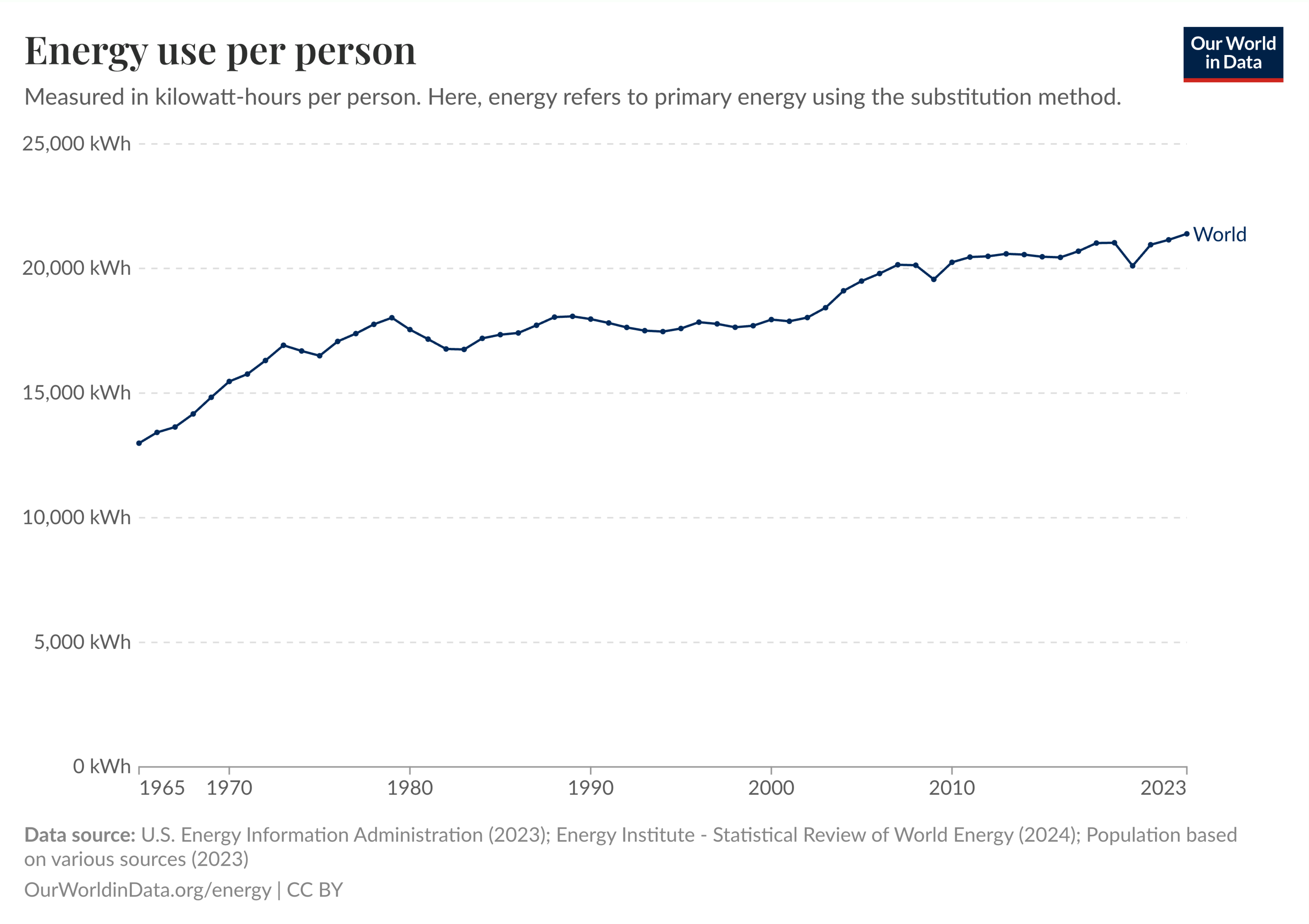
[Accessed September 14], https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2024&locations=EE&start=1993&view=chart
https://www.worldometers.info/gdp/estonia-gdp/

OTHER ADJUSTMENTS (2)

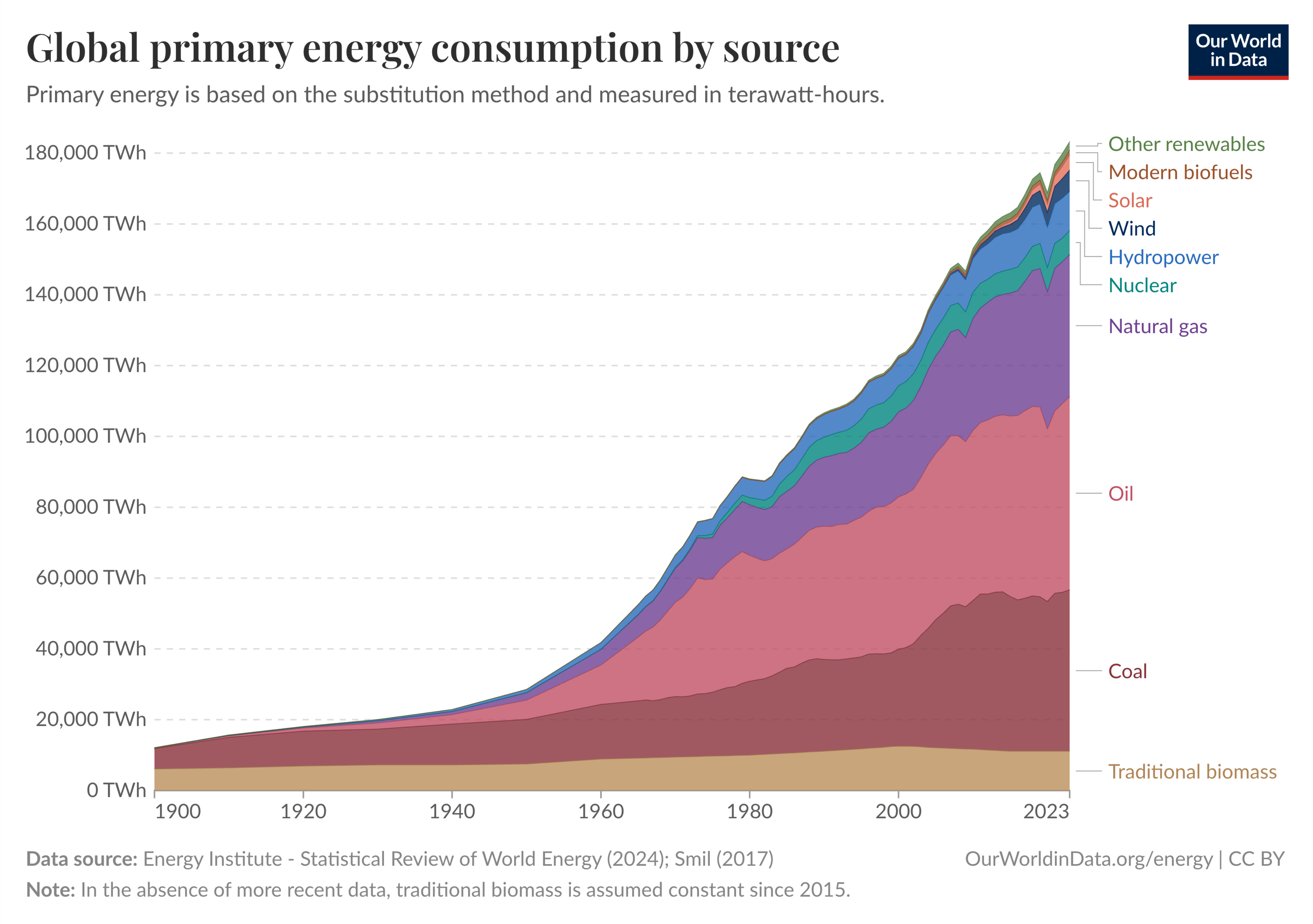


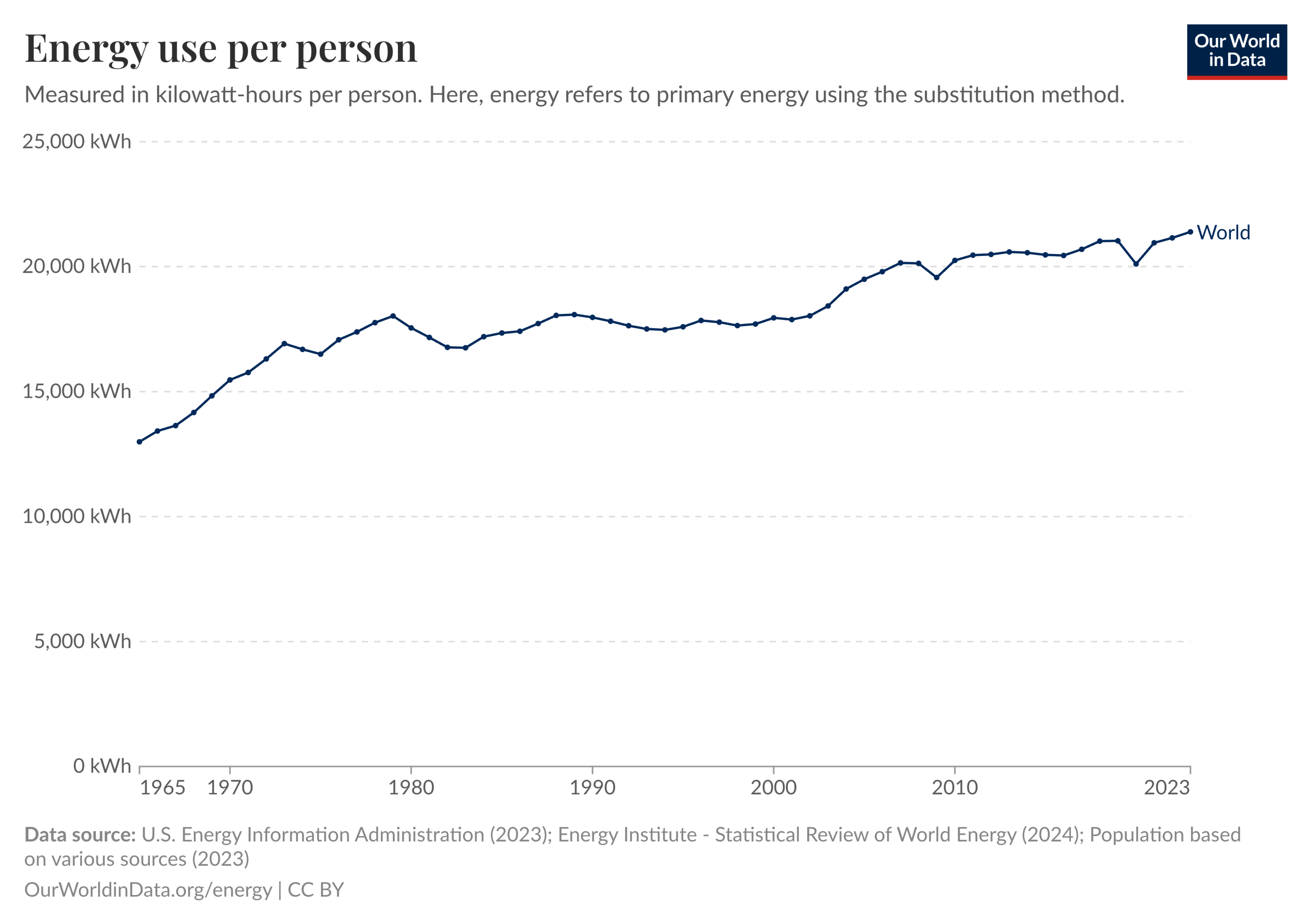
OTHER ADJUSTMENTS (2)





OTHER ADJUSTMENTS (2)





Rule of thumb:

- Add adjusted variable → Creation
- Replace the original variable → Transformation

FEATURE EXTRACTION

Feature extraction is the process of creating new features from existing ones.

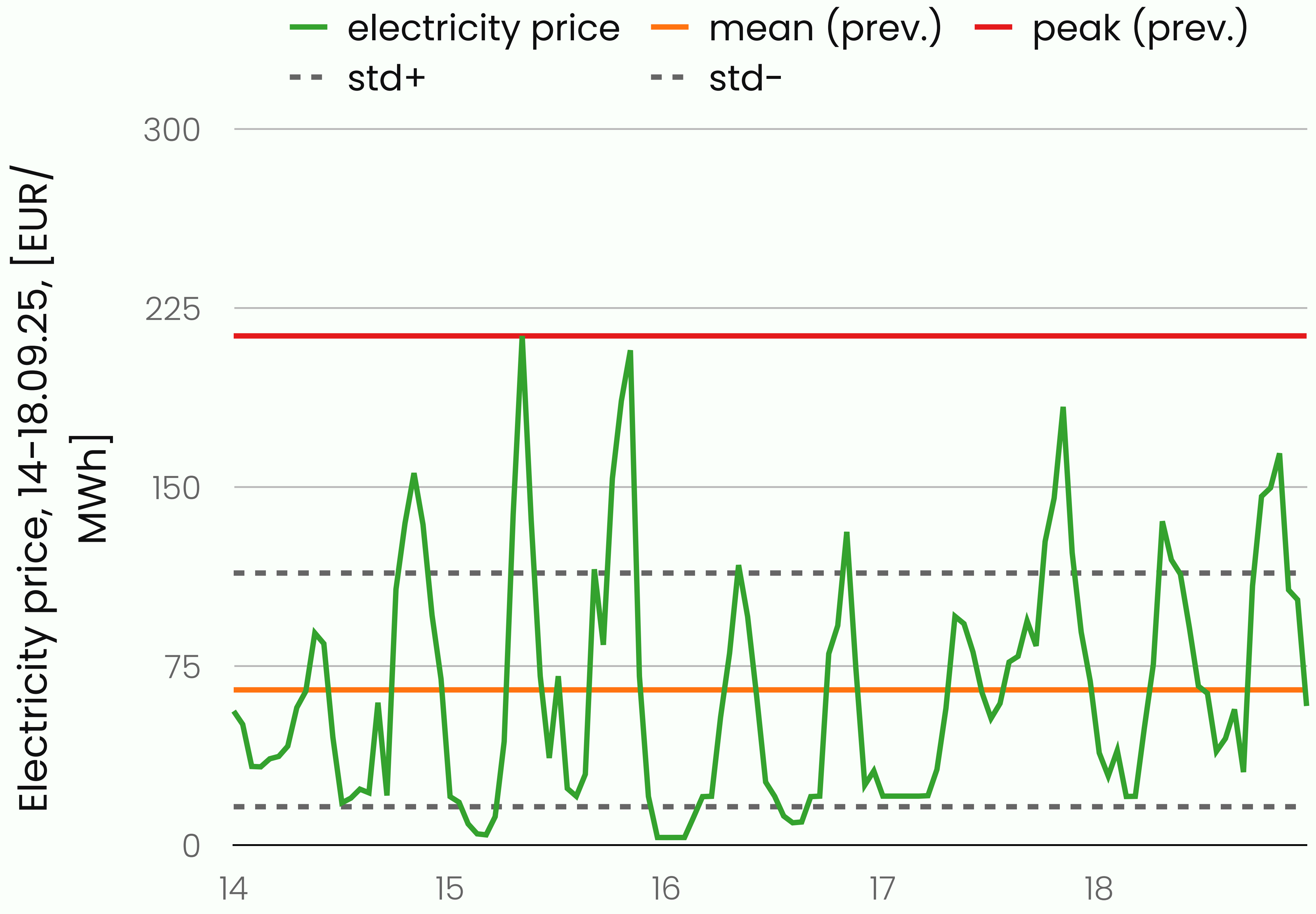
Main goal is to improve:

- Accuracy (by mining relevant patterns)
- Efficiency (by reducing dimensionality)
- Generalisation (by reducing noise)

METHODS

- Numeric data: PCA, LDA, Fourier transform, etc.
- Text: Word embeddings, transformers, bag of words, etc.
- Images: CNNs, etc.
- Time series: descriptive statistics, autoregressive features, etc.

DESCRIPTIVE STATISTICS



Feature	Values
mean / median / mode	65.35 / 57.44 / 20.84
min / max	3.5 / 213.65
var / std	2396.28 / 48.95

PCA VIA SVD

Input:

- Dataset $X \in \mathbb{R}^{n \times p}$

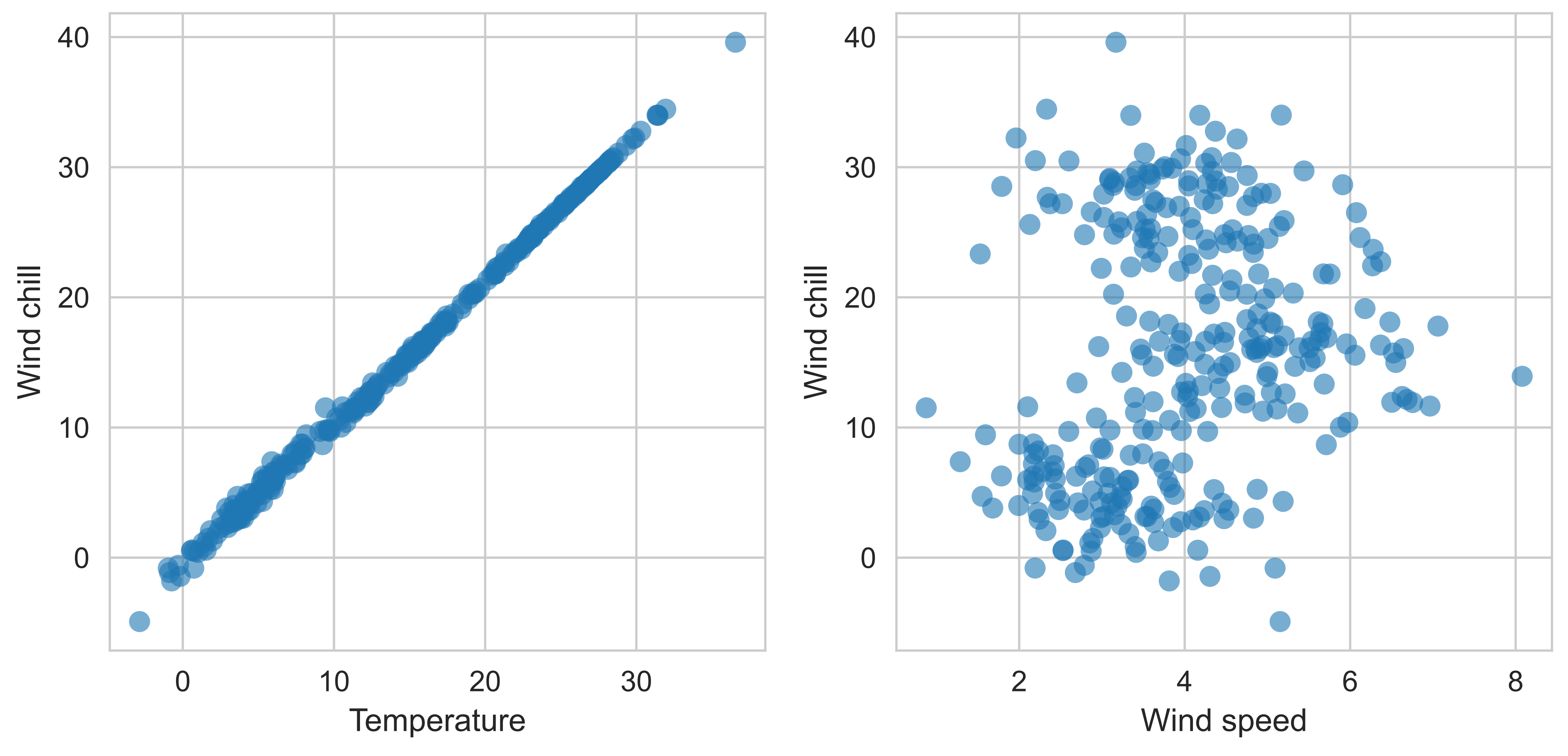
Output:

- Subset of selected features Z in PC space

Algorithm:

1. Center (and *scale*) the data: $\tilde{X} = X - \mu$ ($\tilde{X} = \dfrac{X - \mu}{\sigma}$)
2. SVD decomposition: $\tilde{X} = U \Sigma V^{\top}$
3. Principal axis (loadings): $V = [v_1, \ldots, v_p]$
4. Principal components: $Z = U\Sigma$
5. Explained variance ratio (for PCi): $\mathrm{EVR}_i = \dfrac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$

PCA



Target variable: demand

Features:

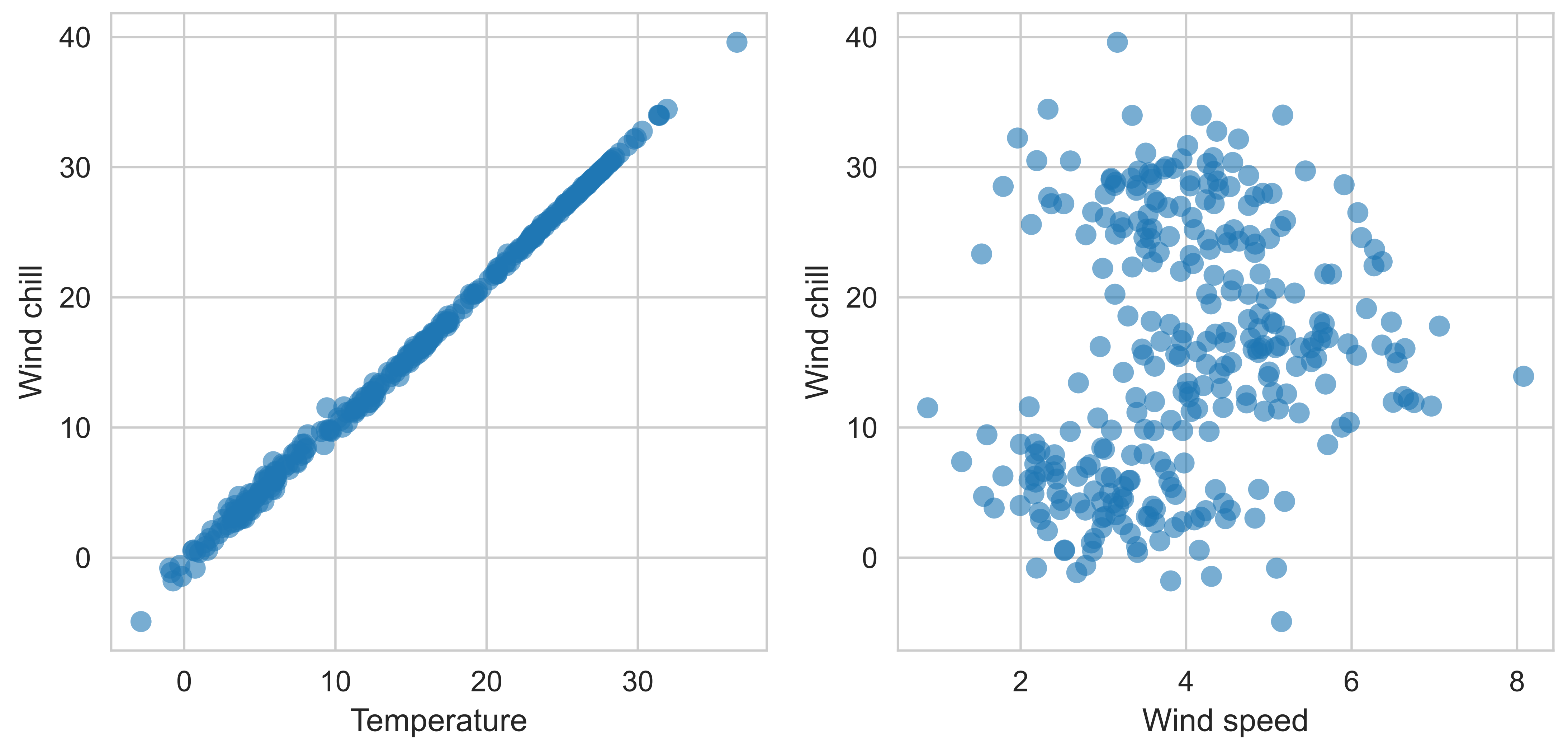
- Temperature
- Wind speed
- Wind chill

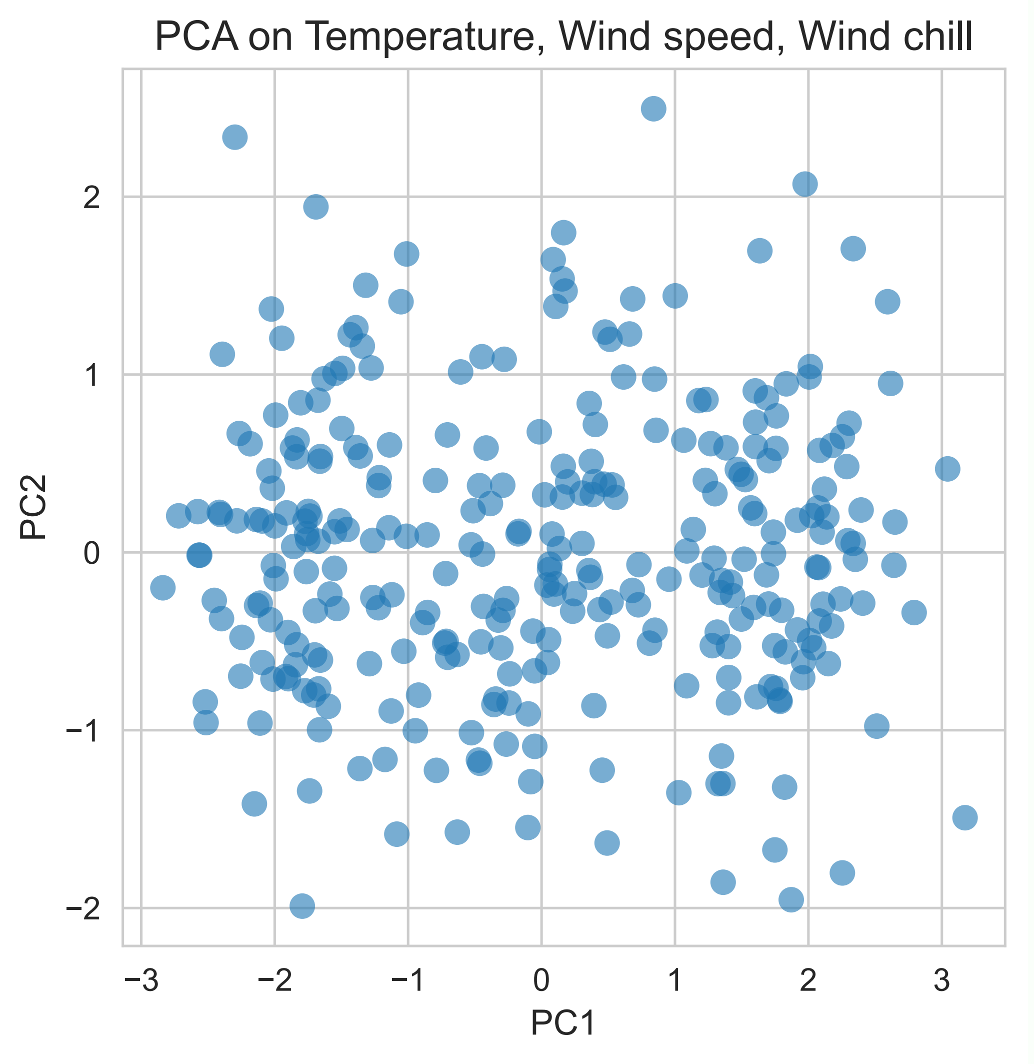
PCA

Target variable: demand

Features:

- Temperature
- Wind speed
- Wind chill



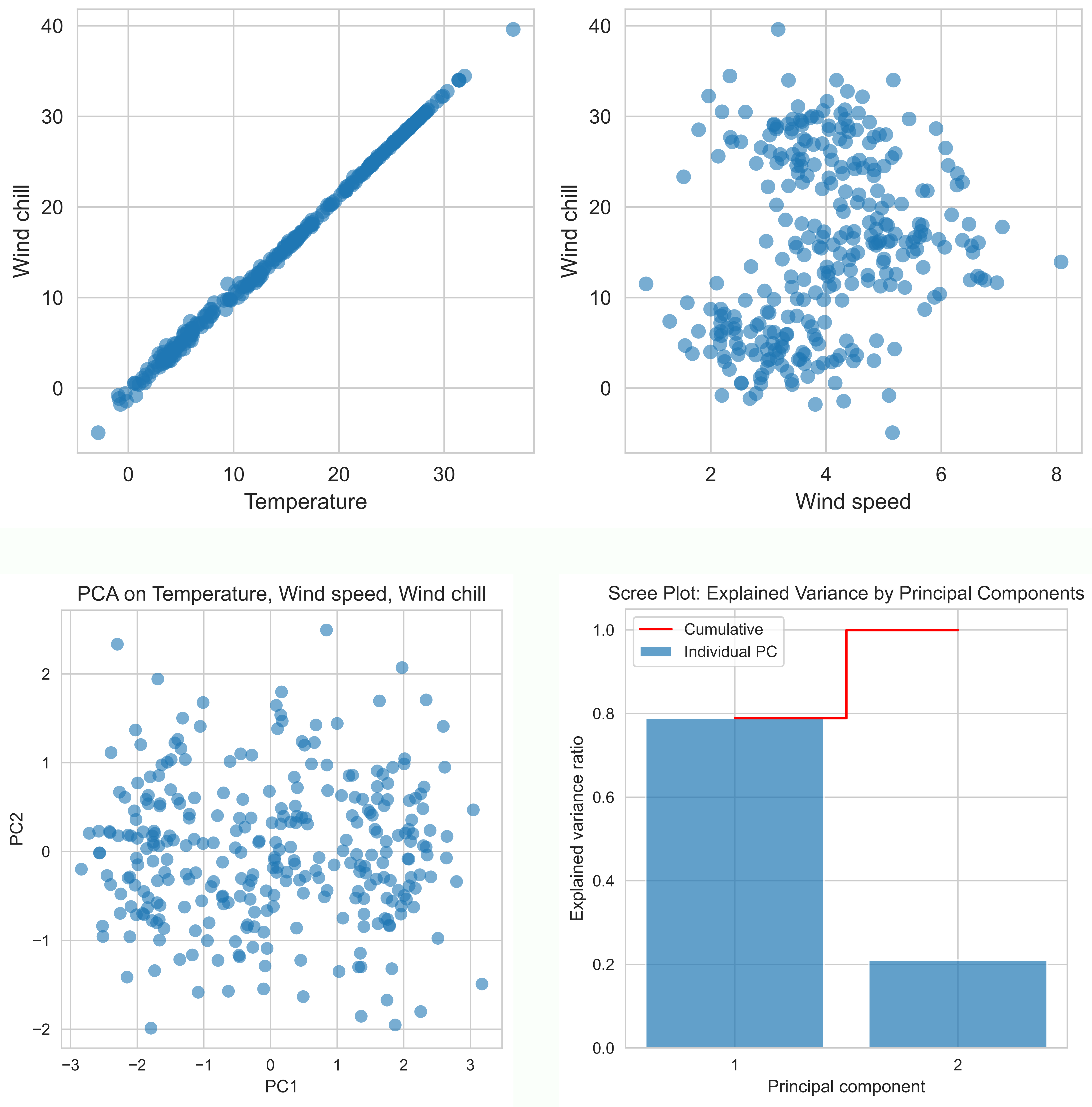


PCA

Target variable: demand

Features:

- Temperature
- Wind speed
- Wind chill



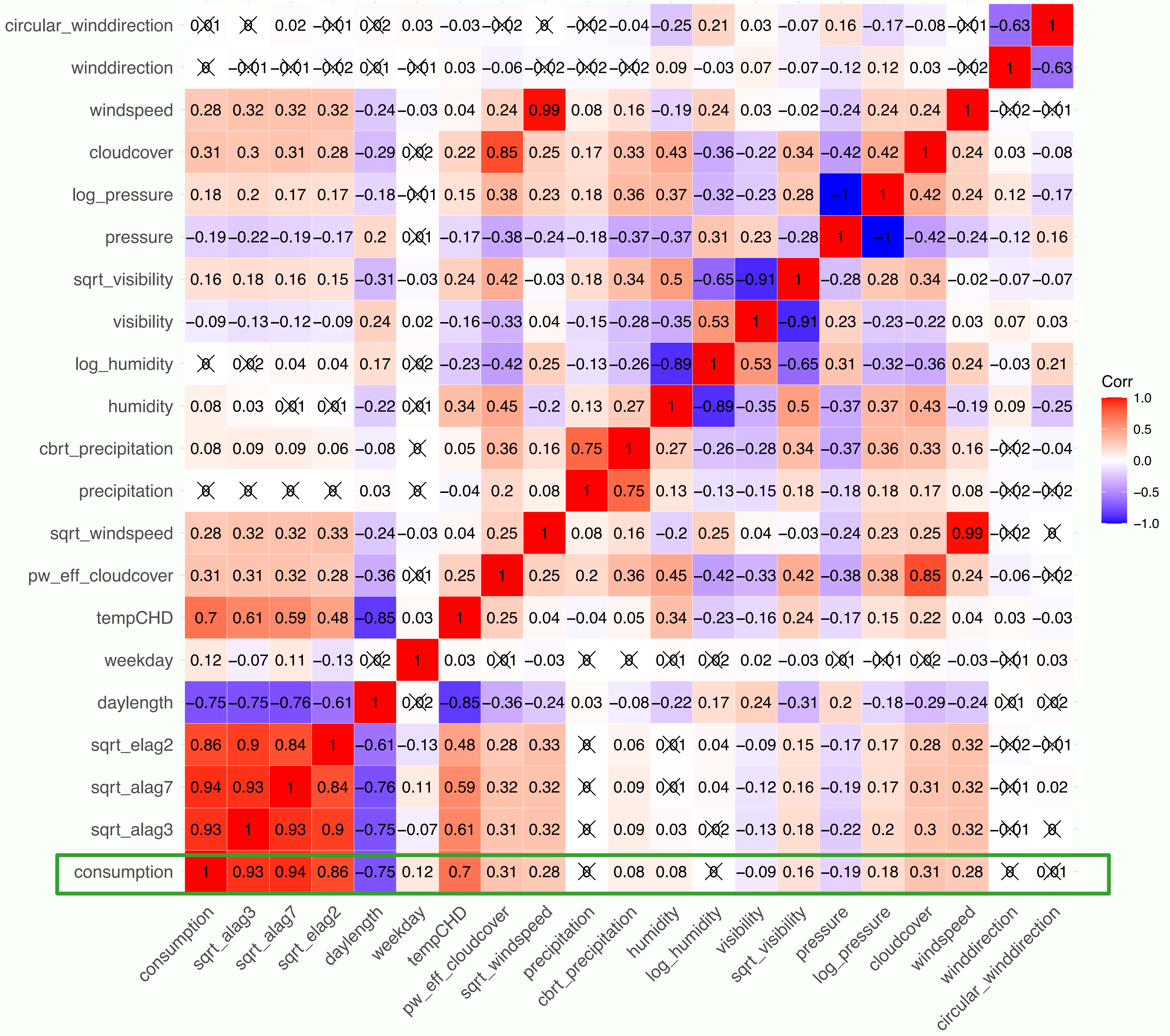
FEATURE SELECTION

Feature selection is the process of choosing a subset of important features from a dataset.

Common methods:

- Filter methods (correlations, Chi-squared test, mutual information, etc.)
- Wrapper methods (forward selection, backward elimination, etc.)
- Embedded methods (LASSO, tree-based, etc.)

PEARSON CORRELATION



M. Sinimaa, M. Spichakova, J. Belikov, E. Petlenkov. Feature engineering of weather data for short-term energy consumption forecast. *IEEE Madrid PowerTech*: *Madrid, Spain*, 2021.

FORWARD SELECTION (BACKWARD ELIMINATION)

SUMMARY OF VARIABLES USED IN VARIOUS MODELS

Var.	0a	0b	1a	2a	3a	4a	4b	5a	5b	6a	7a	7b	8a	8b
y	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
$\sqrt{y}$													✓	✓
x_1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
x_2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
x_3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
x_4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
x_5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
x_6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
x_7			✓											
x_8										✓		✓	✓	✓
x_9					✓									
x_{10}				✓										
x_{11}									✓		✓	✓	✓	✓
x_{12}								✓						
x_{13}						✓								
x_{14}							✓							
x_{15}													✓	✓
x_{16}													✓	✓
x_{17}													✓	✓

Input:

- Dataset X with features $F = \{f_1, f_2, \dots, f_n\}$
- Target variable y
- Evaluation criterion
- Stopping rule

Output:

- Subset of selected features S

Algorithm:

1. Initialise $S = \varnothing$ ($S = F$)
2. Repeat until stopping rule is met:
   (a) For each feature $f \in f \backslash S$ ($f \in S$):
      - Train a model using $S \cup \{f\}$ ($S \backslash \{f\}$)
      - Evaluate its performance using the criterion

   (b) Select the feature f^* whose addition (removal) improves performance

   (c) Update $S \leftarrow S \cup \{f^*\}$ ($S \leftarrow S \backslash \{f^*\}$)

   (d) If no feature improves the performance → stop
3. Return the final subset S

M. Sinimaa, M. Spichakova, J. Belikov, E. Petlenkov. Feature engineering of weather data for short-term energy consumption forecast. *IEEE Madrid PowerTech: Madrid, Spain*, 2021.

FEATURE ENGINEERING TOOLS

Library	Feature engineering	Feature selection	Open source	Support for time series
Scikit-learn	Yes	Yes	Yes	No
Feature Engine	Yes	Yes	Yes	Yes
Featuretools	Yes	Ye	Yes	Yes
AutoFeat	Yes	Yes	Yes	No
TSFresh	Yes	Yes	Yes	Yes

HOME ACTIVITIES & BRAIN EXERCISE

None.

Work *hard* on your project :).

Thank you!

Questions?