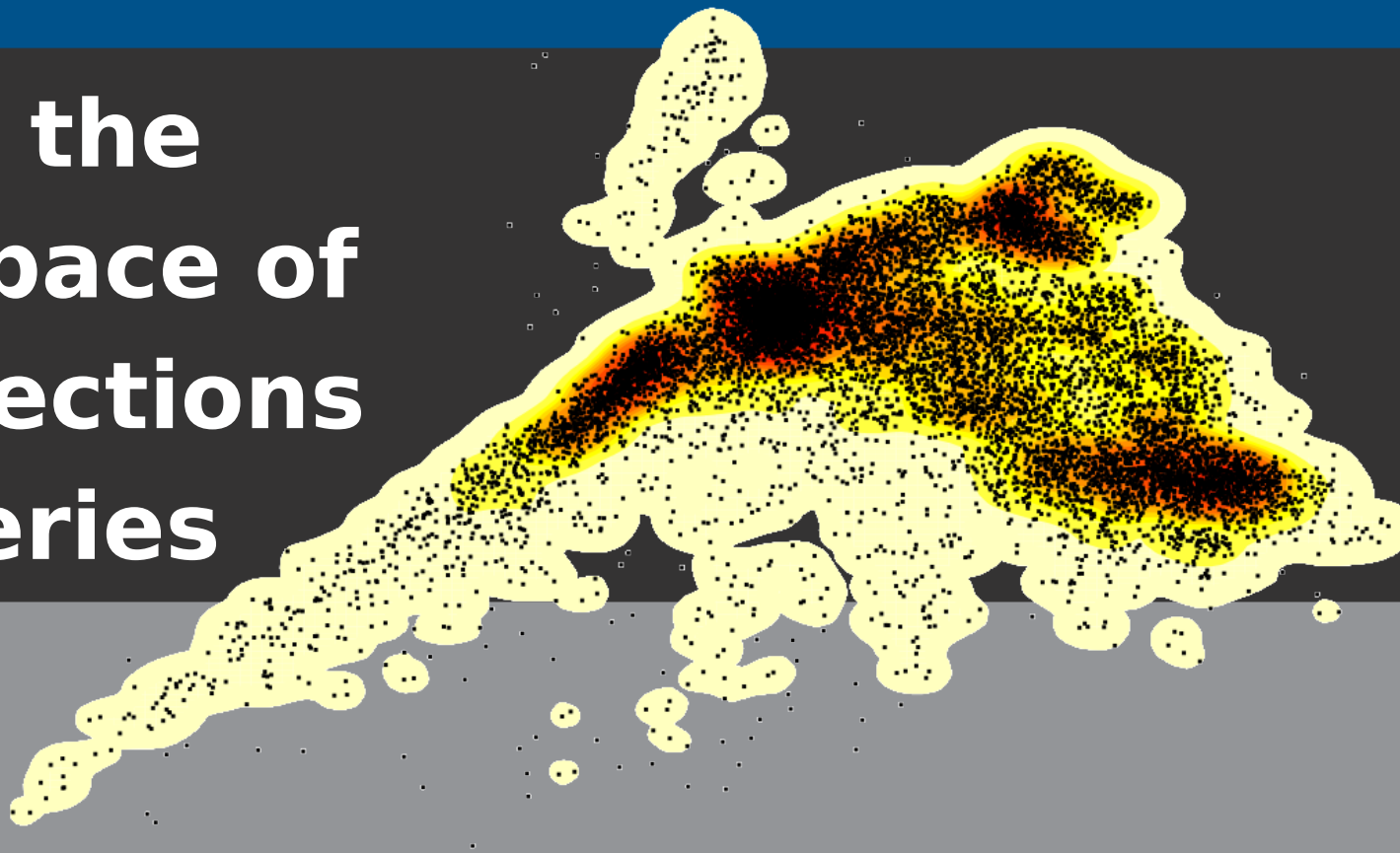
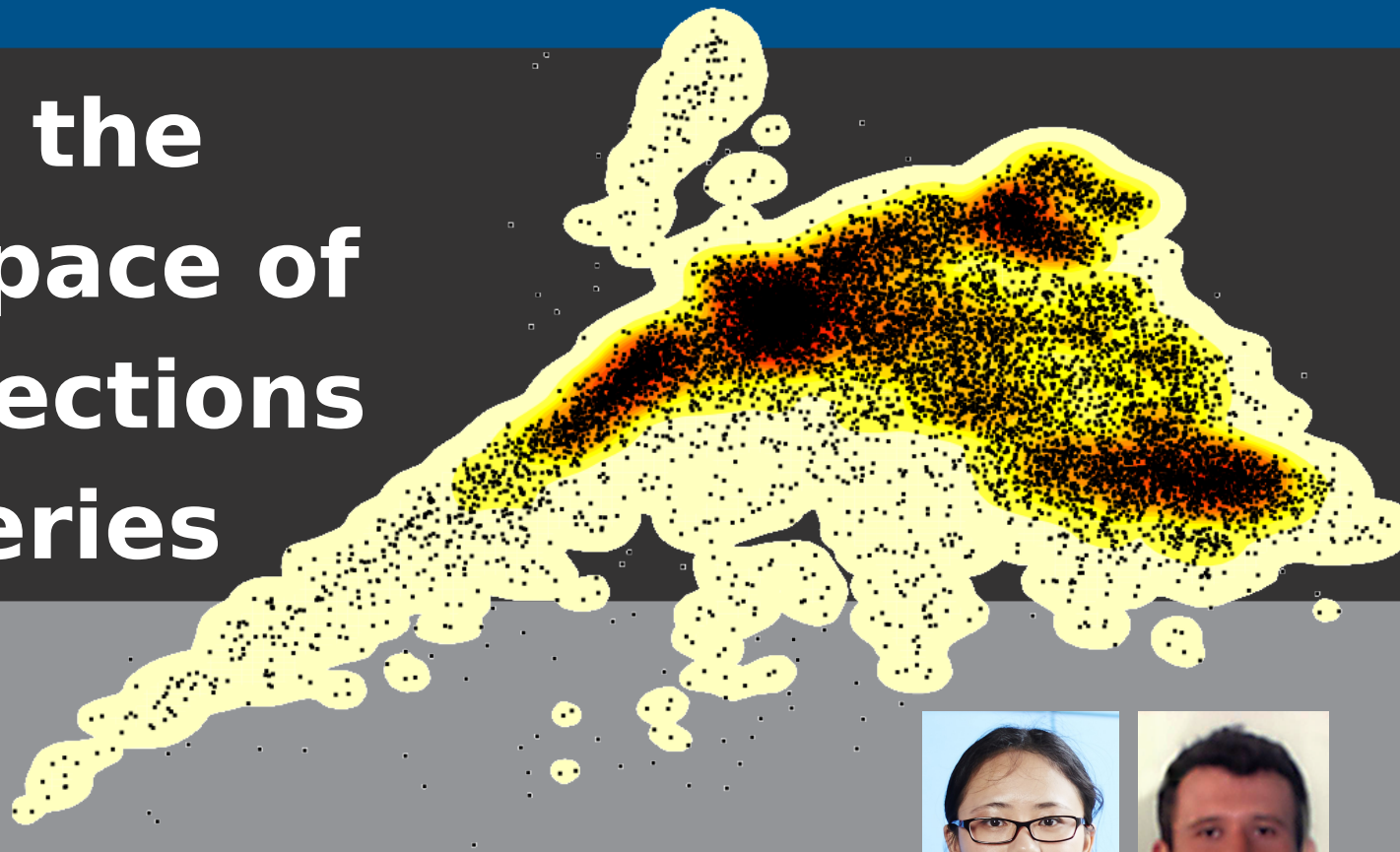
Exploring the feature space of large collections of time series



Rob J Hyndman

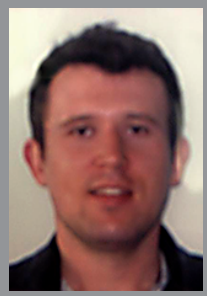
MONASH BUSINESS SCHOOL

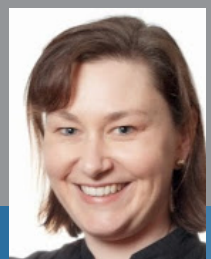
Exploring the feature space of large collections of time series



Rob J Hyndman

with Earo Wang, Nikolay Laptev
Yanfei Kang, Kate Smith-Miles





Outline

M3 forecasting competition

ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

international journal of forecasting

www.elsevier.com/locate/ijforecast

The M3-Competition: results, conclusions and implications

Spyros Makridakis, Michèle Hibon*

INSEAD, Boulevard de Constance, 77305 Fontainebleau, France

Abstract

This paper describes the M3-Competition, the latest of the M-Competitions. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results/conclusions with those of the previous two M-Competitions as well as with those of other major empirical studies. Finally, the implications of these results and conclusions are considered, their consequences for both the theory and practice of forecasting are explored and directions for future research are contemplated.

Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting accuracy

M3 forecasting competition

ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

international journal of forecasting

www.elsevier.com/locate/ijforecast

results, conclusion s

Makridakis, Michèle Hibo

ard de Constance, 77305 Fontainebleau

Abstract

This pa he latest of the M-Competitions. I ucting the competiti clusions. In addition, the paper com with those of the pre with those of other major empirical s of these results and conclusions are considered, their consequences for both the theory and practice of forecasting are explored and directions for future research are contemplated.

Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting accuracy

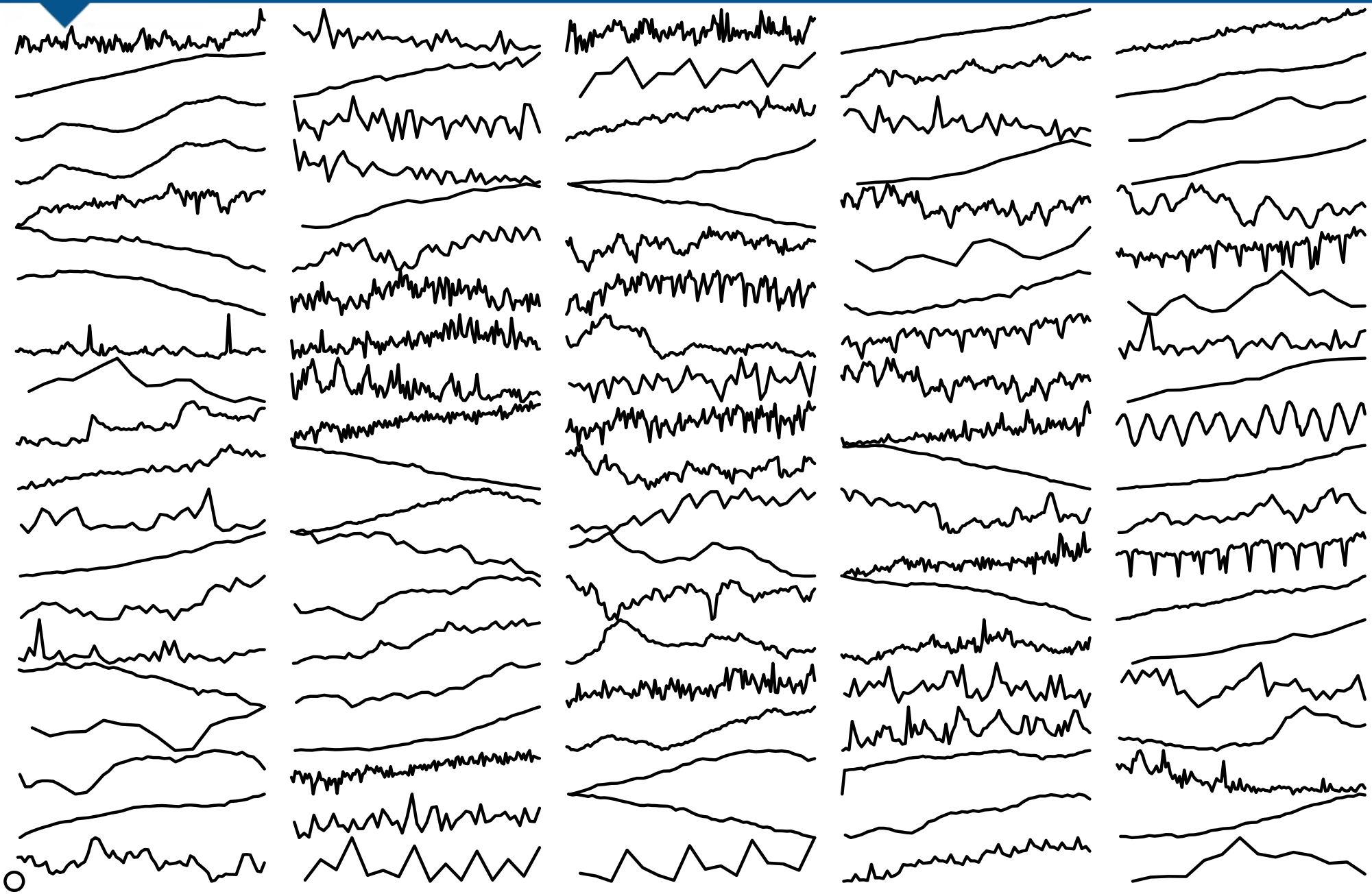
M3 forecasting competition

"The M3-Competition is a final attempt by the authors to settle the accuracy issue of various time series methods. . . The extension involves the inclusion of more methods/ researchers (in particular in the areas of neural networks and expert systems) and more series."

Makridakis & Hibon, IJF 2000

- 3003 series
- All data from business, demography, finance and economics.
- Series length between 14 and 126.
- Either non-seasonal, monthly or quarterly.
- All time series positive.

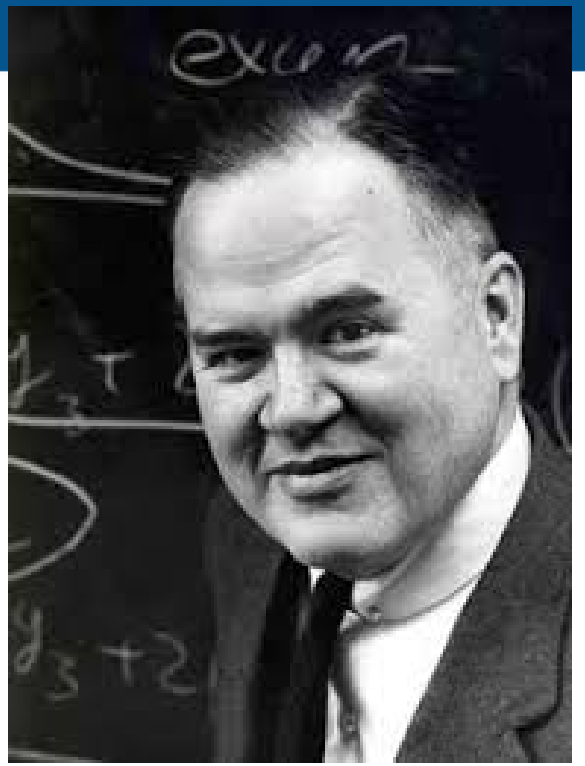
M3 forecasting competition



Key idea

Cognostics

Computer-produced diagnostics (Tukey and Tukey, 1985).



John W Tukey

Examples for time series

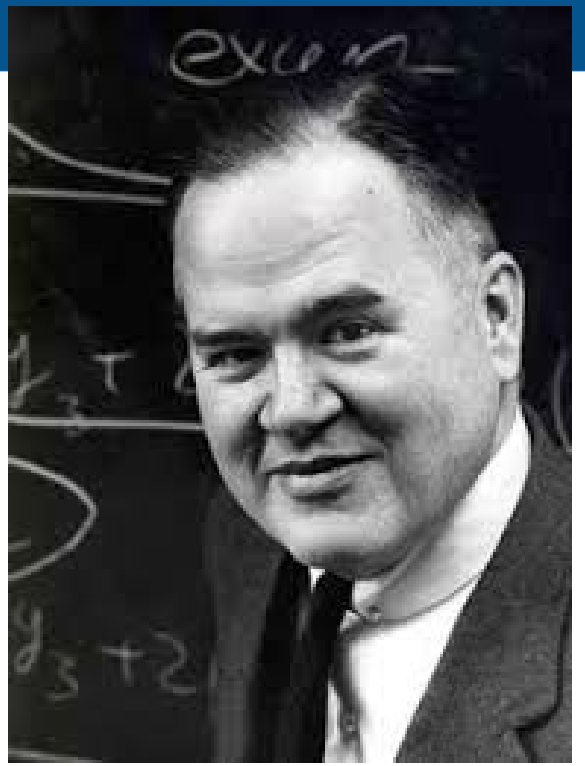
- lag correlation
- size and direction of trend
- strength of seasonality
- timing of peak seasonality
- spectral entropy

Called “features” or “characteristics” in the machine learning literature.

Key idea

Cognostics

Computer-produced diagnostics (Tukey and Tukey, 1985).



John W Tukey

Examples for time series

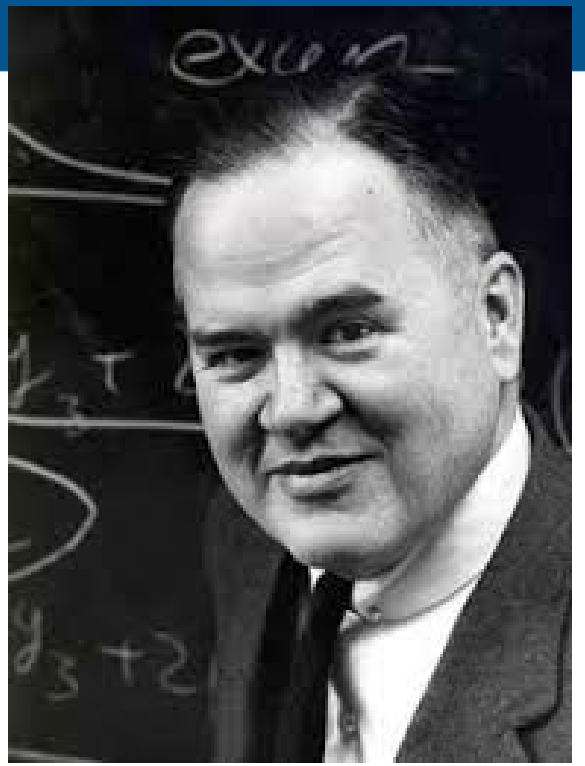
- lag correlation
- size and direction of trend
- strength of seasonality
- timing of peak seasonality
- spectral entropy

Called “features” or “characteristics” in the machine learning literature.

Key idea

Cognostics

Computer-produced diagnostics (Tukey and Tukey, 1985).



John W Tukey

Examples for time series

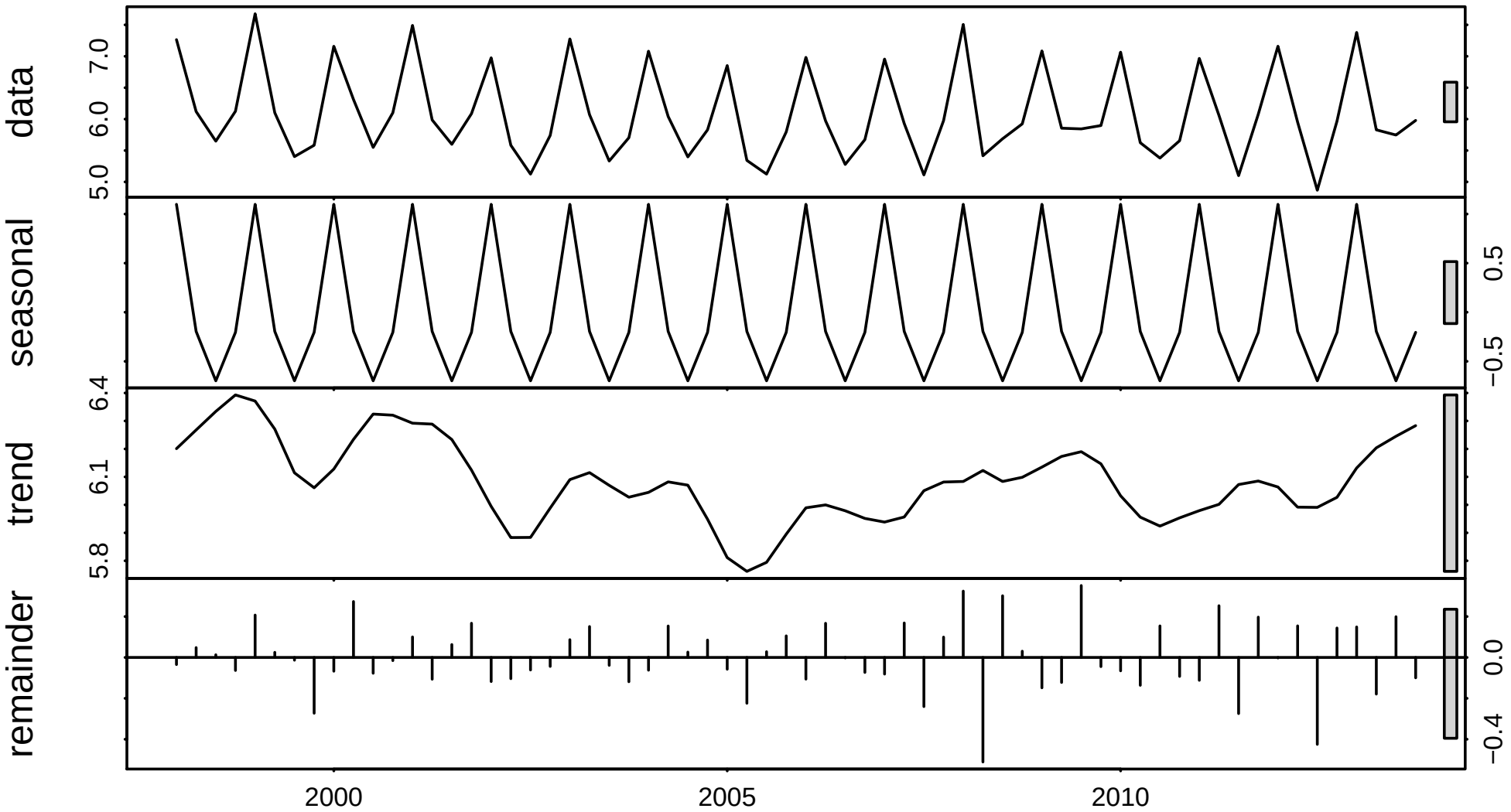
- lag correlation
- size and direction of trend
- strength of seasonality
- timing of peak seasonality
- spectral entropy

Called “features” or “characteristics” in the machine learning literature.

An STL decomposition

Quarterly visitor nights: Mornington Peninsula

$Y_t = S_t + T_t + R_t$ S_t is periodic with mean 0



Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f(\lambda) \log f(\lambda) d\lambda$, where $f(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

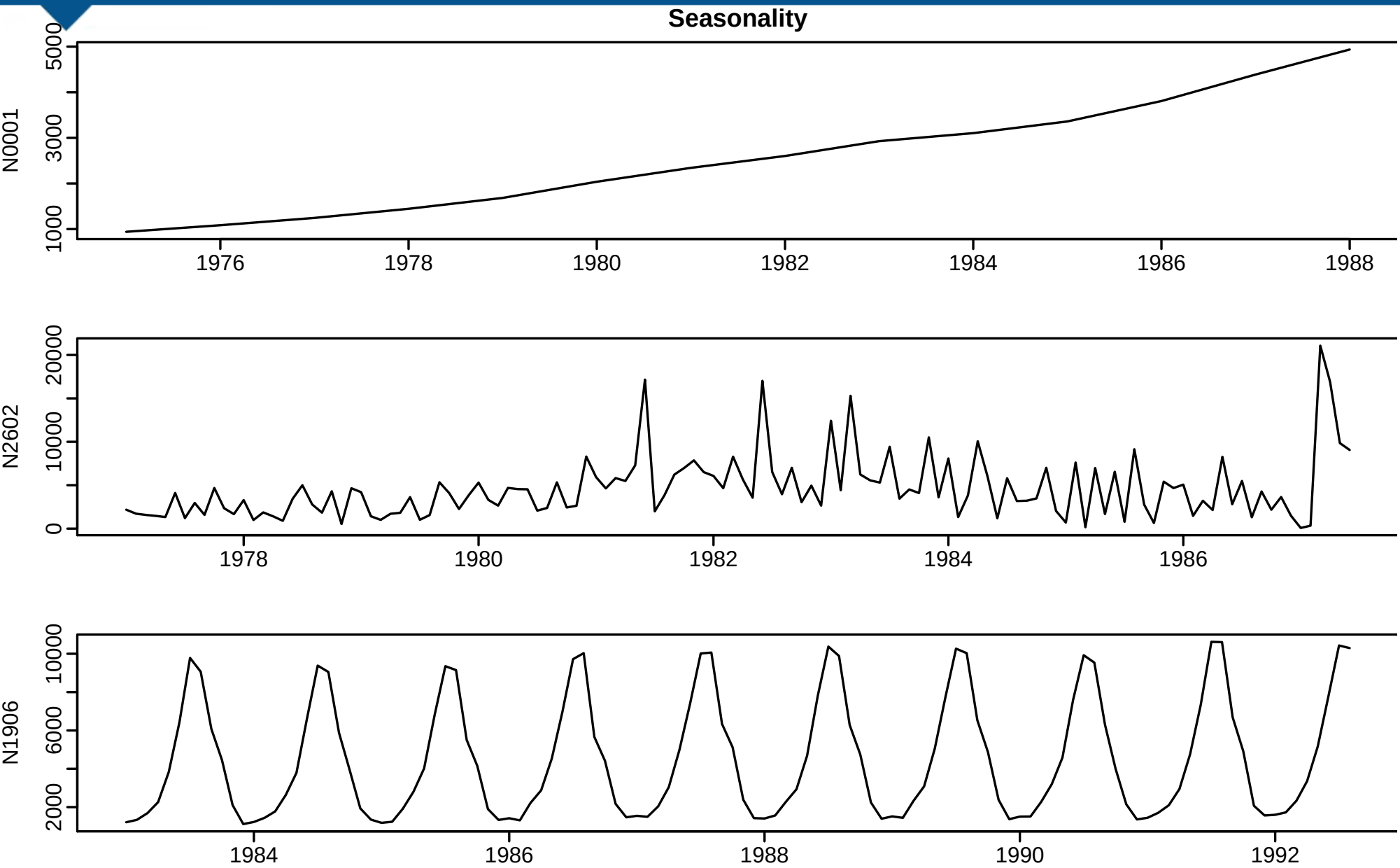
Candidate features

STL decomposition

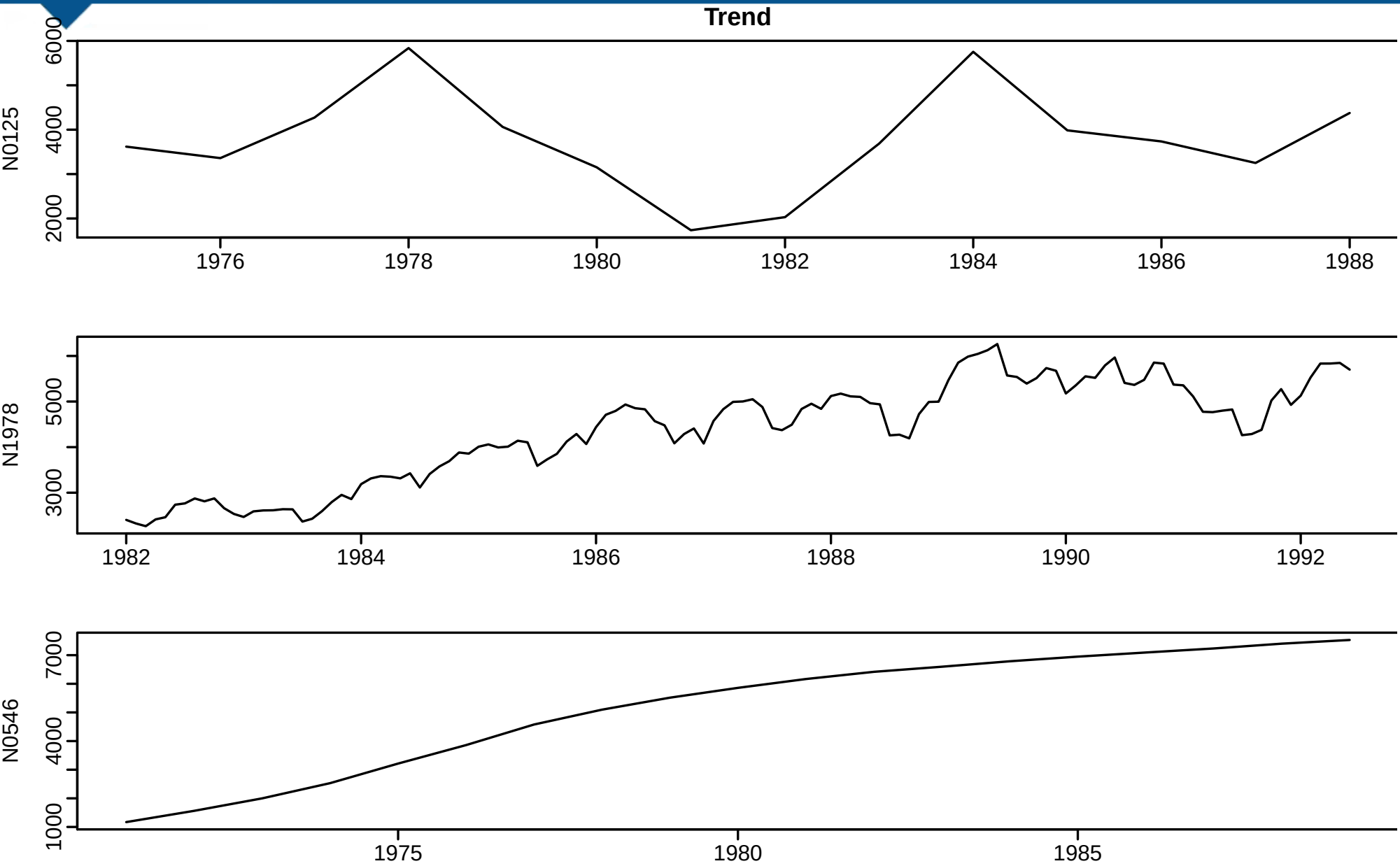
$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Strength of seasonality: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}$
- Strength of trend: $1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Autocorrelations: $r_1, r_2, r_3, \ldots$
- Optimal Box-Cox transformation parameter λ

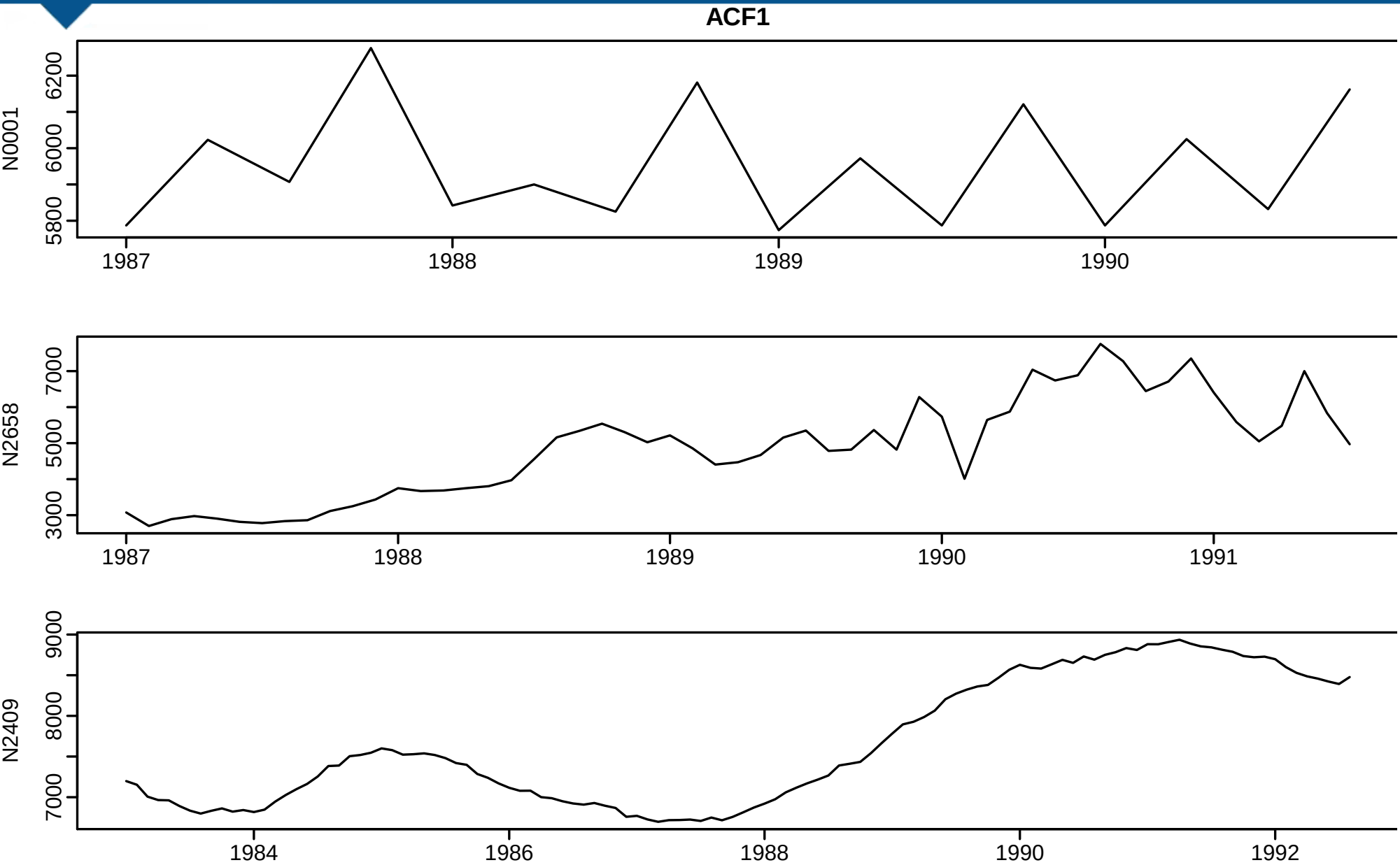
Candidate features



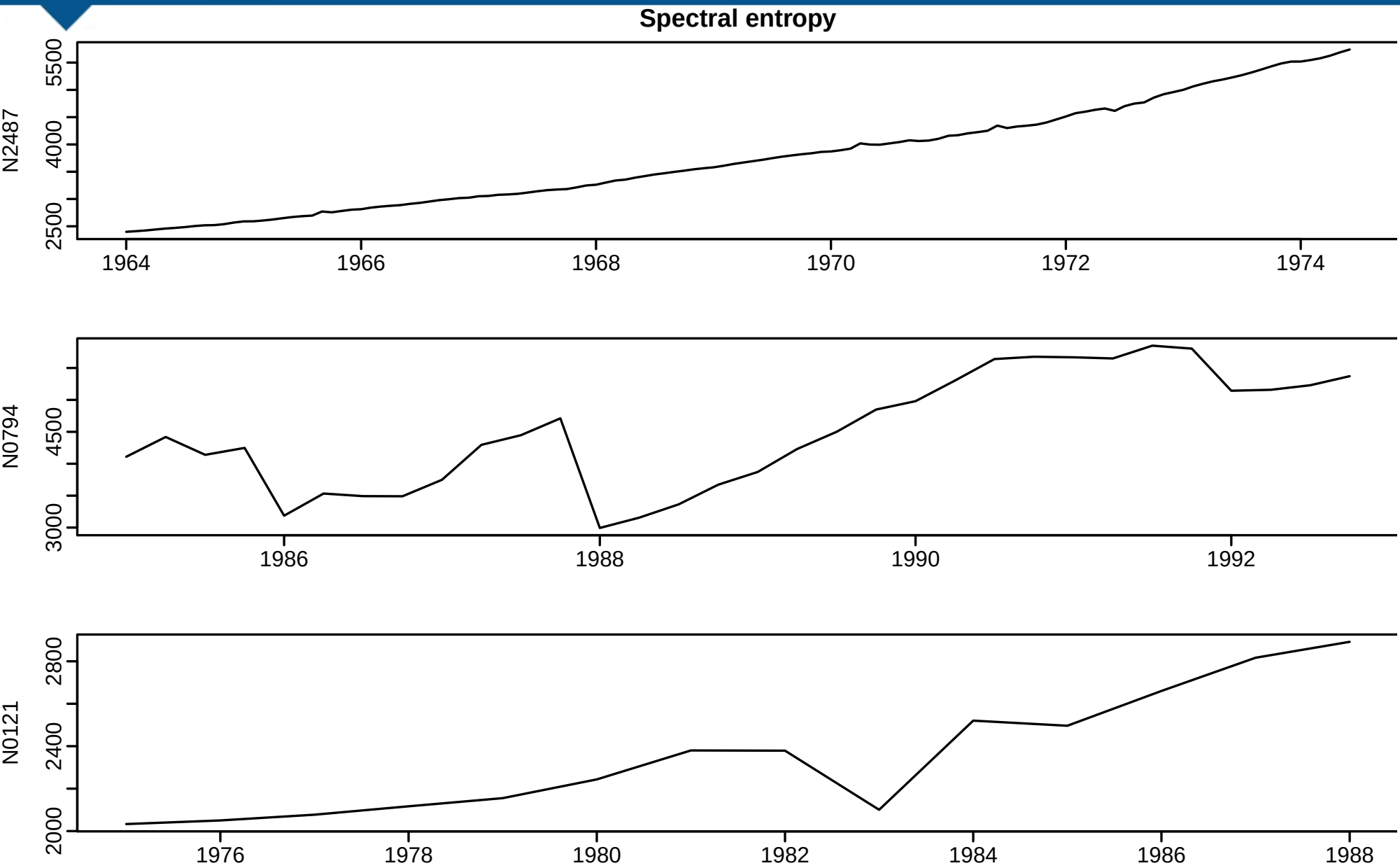
Candidate features



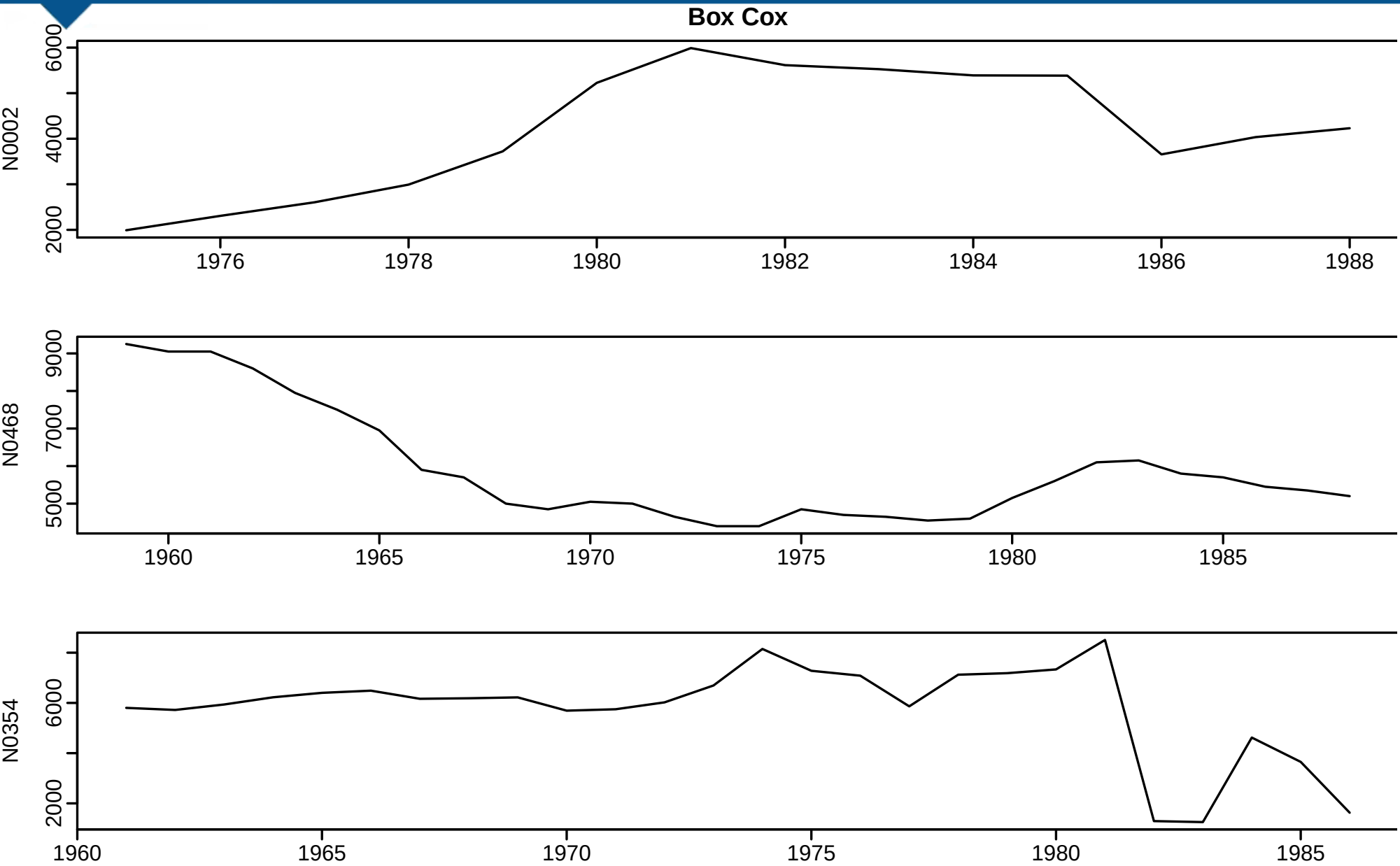
Candidate features



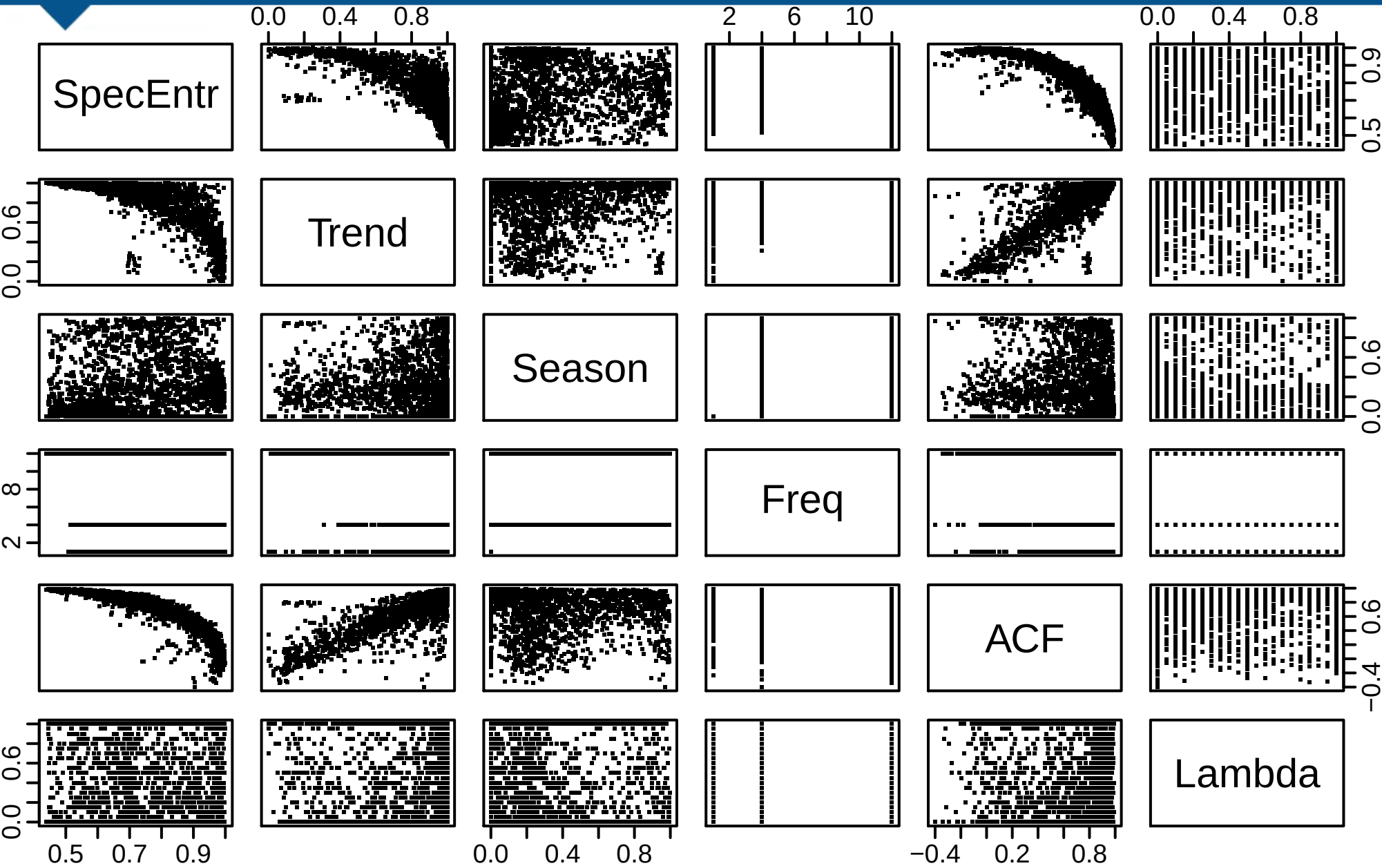
Candidate features



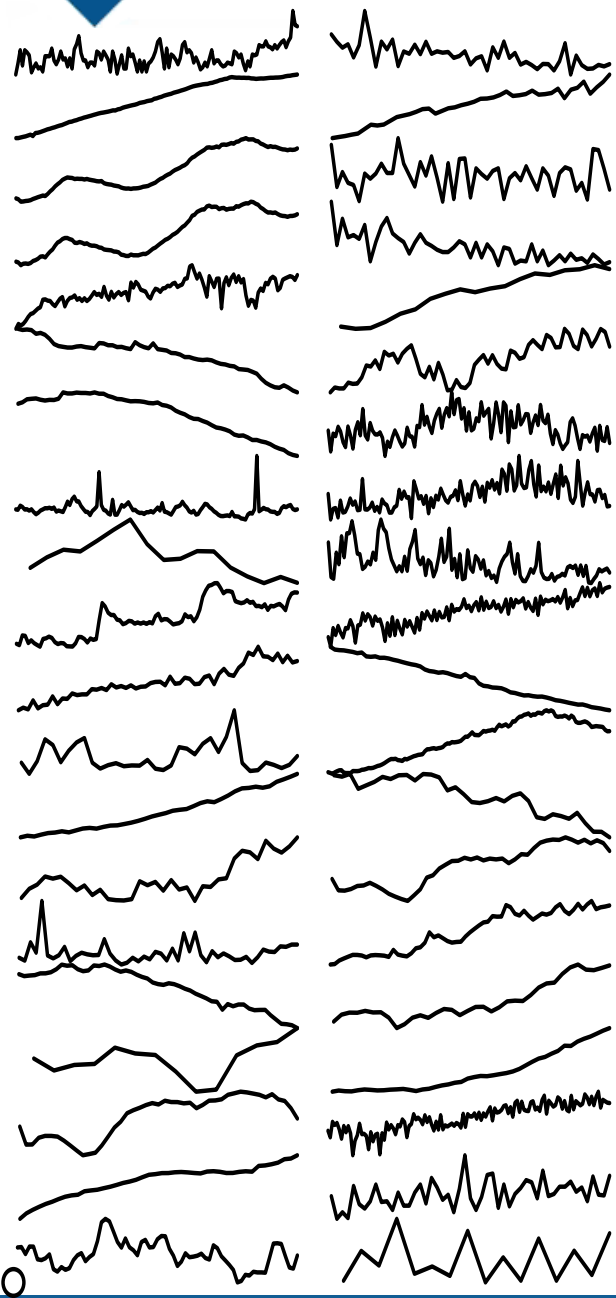
Candidate features



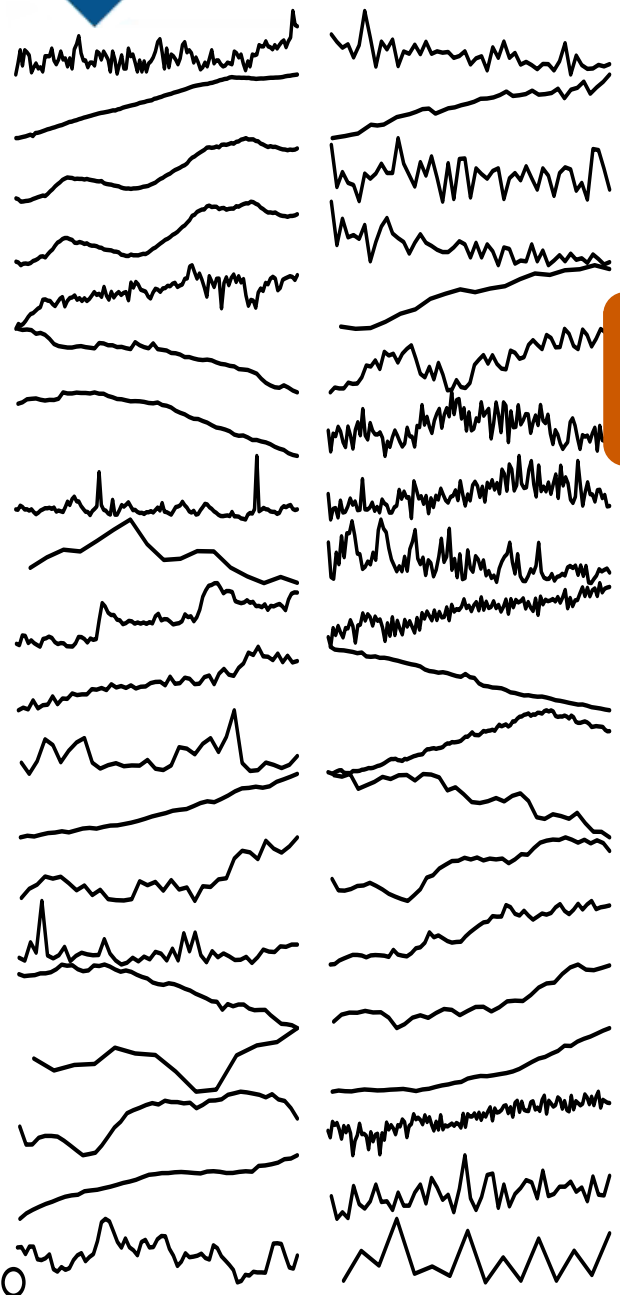
Candidate features



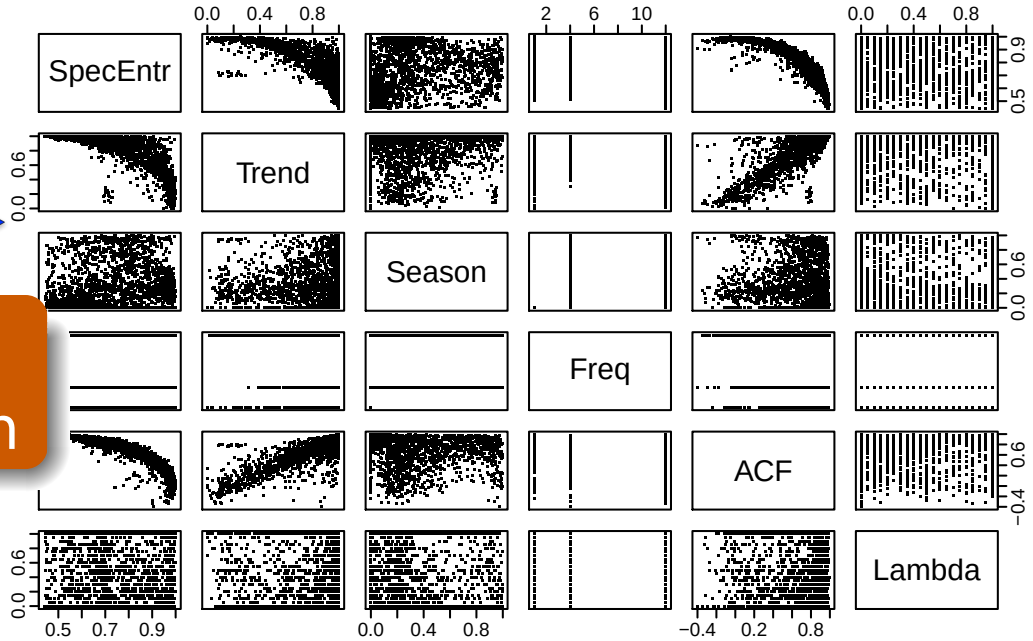
Dimension reduction for time series



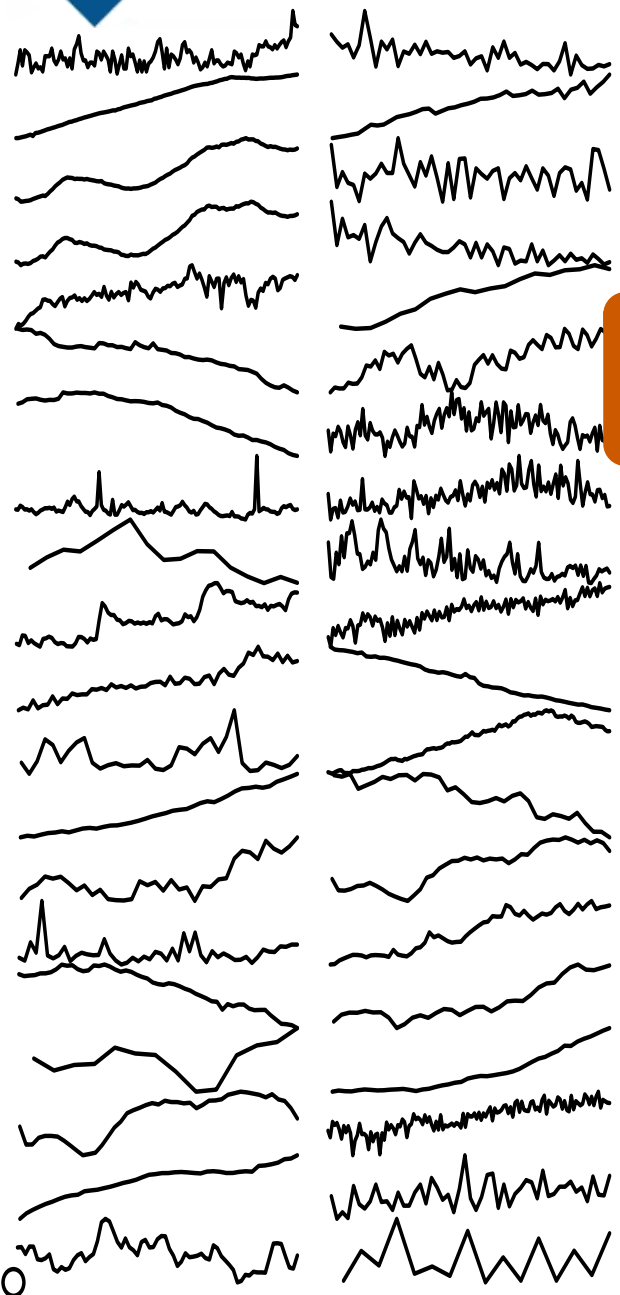
Dimension reduction for time series



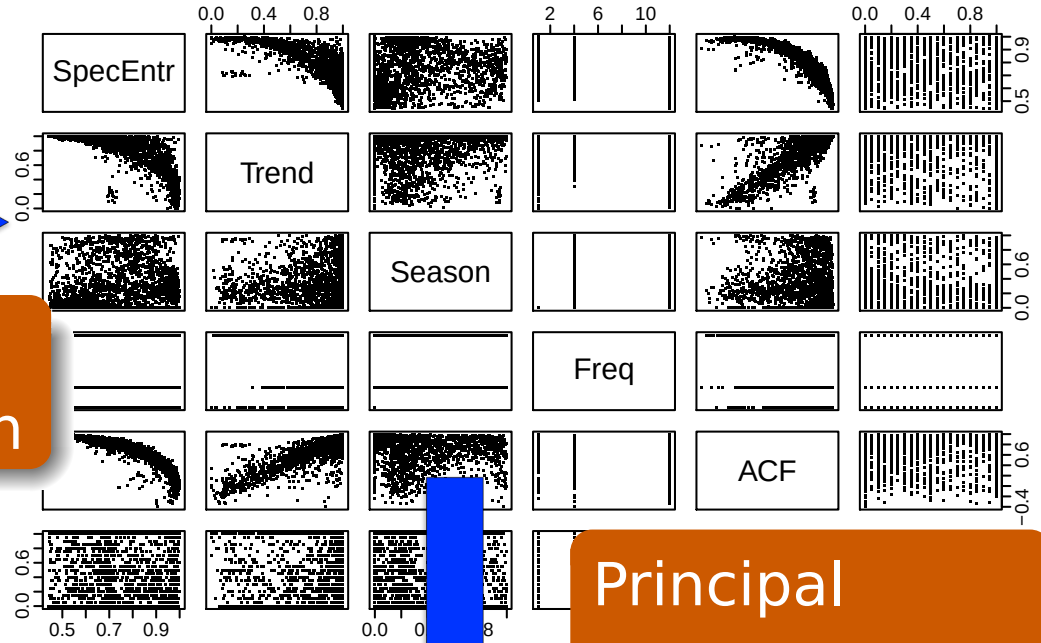
Feature calculation



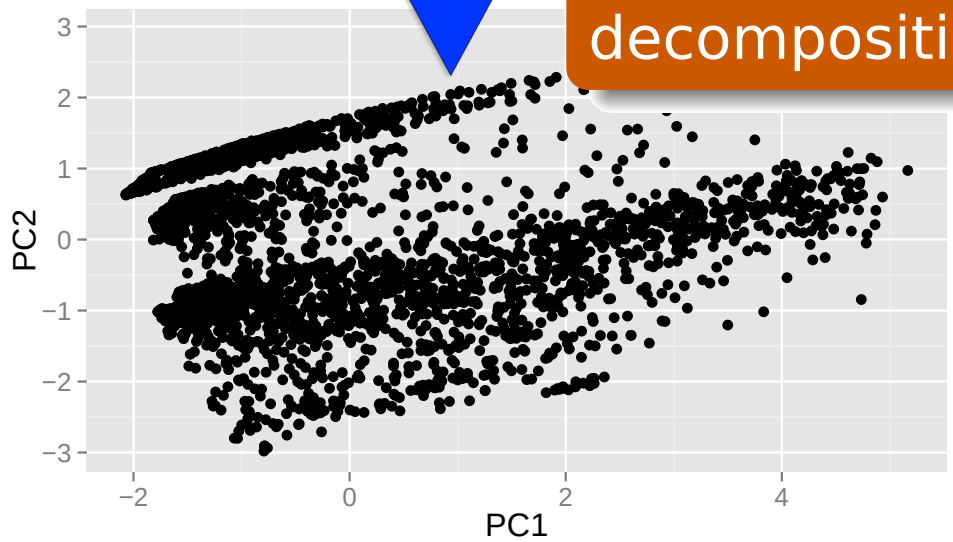
Dimension reduction for time series



Feature calculation

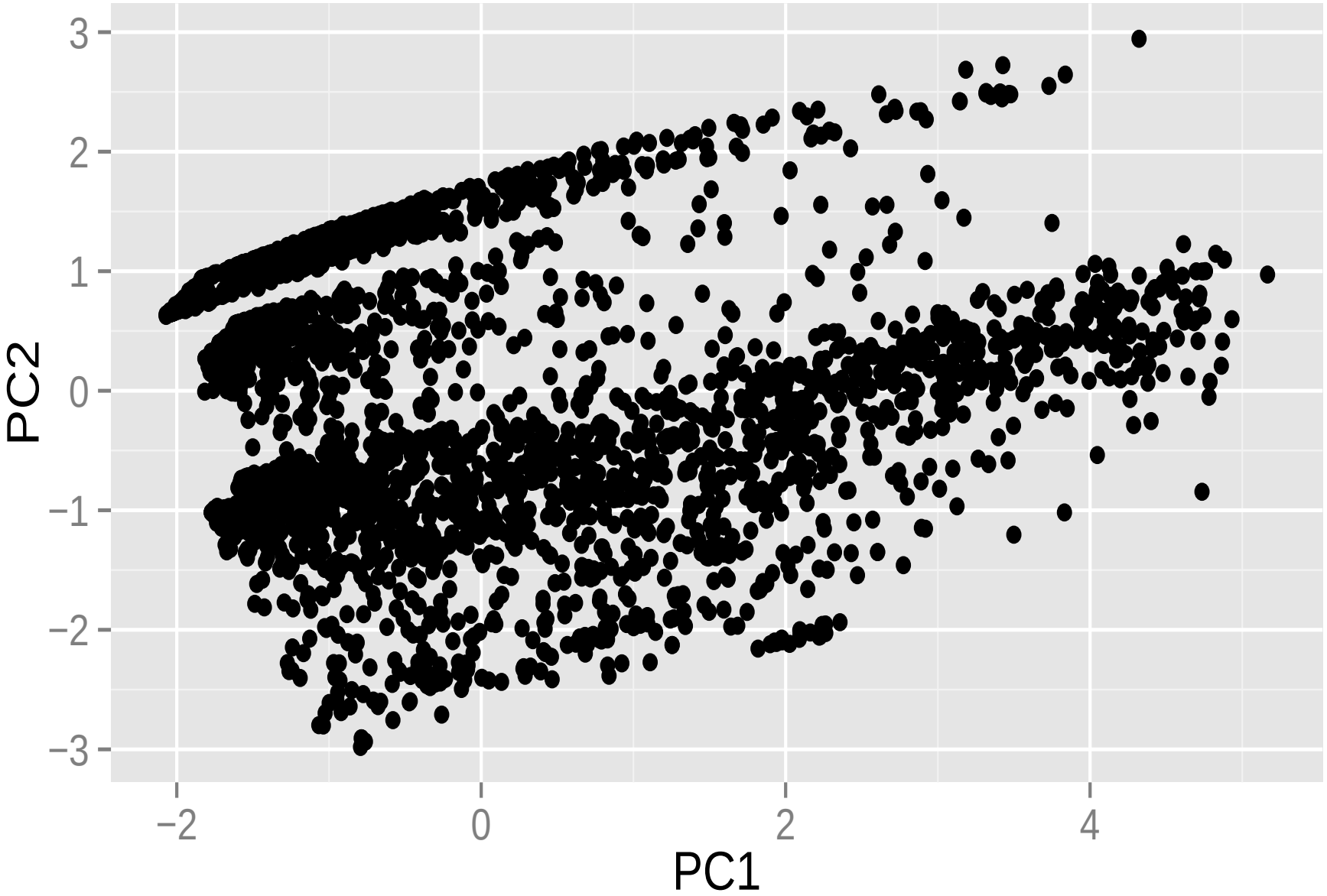


Principal component decomposition

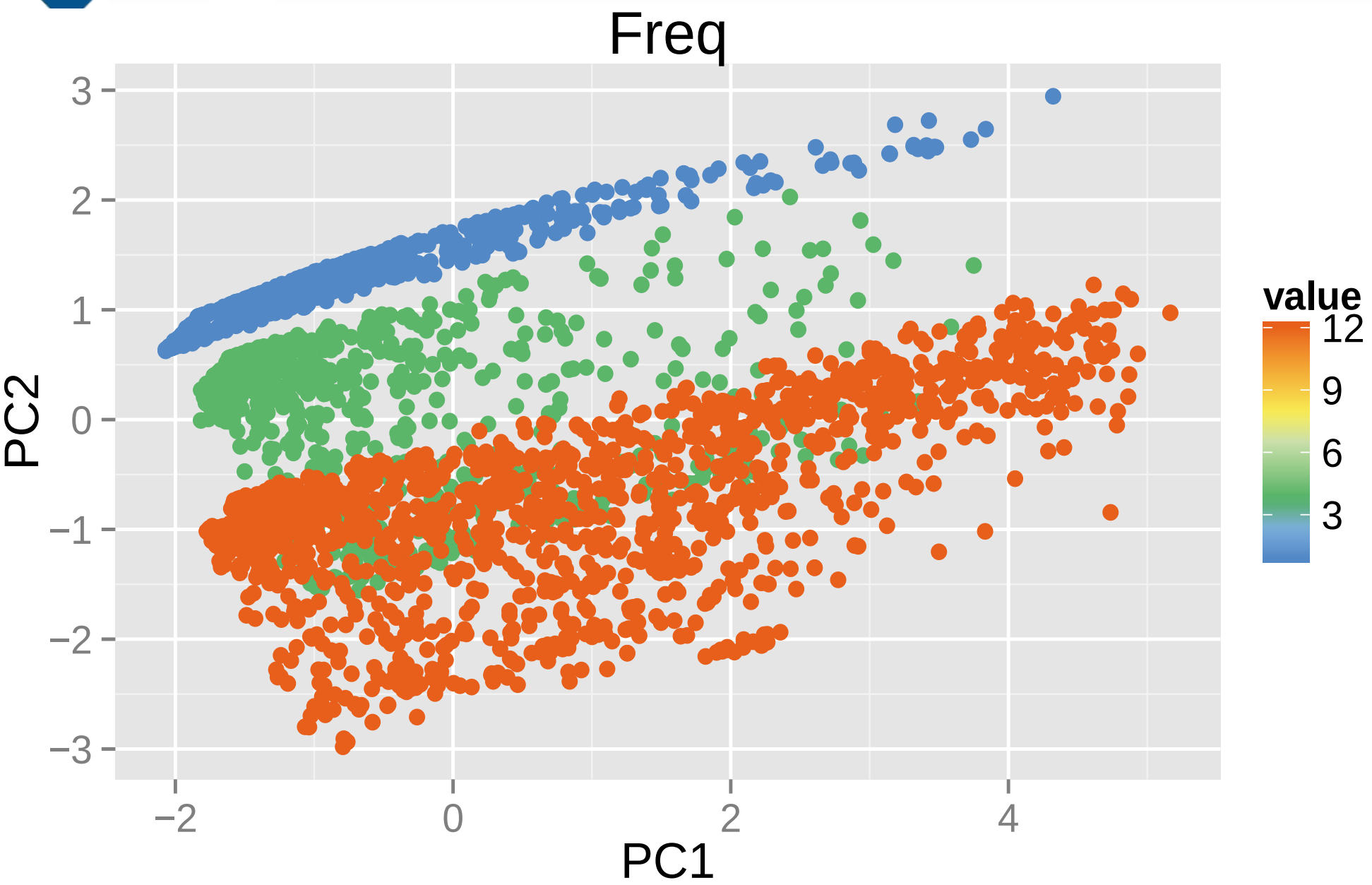


Feature space of M3 data

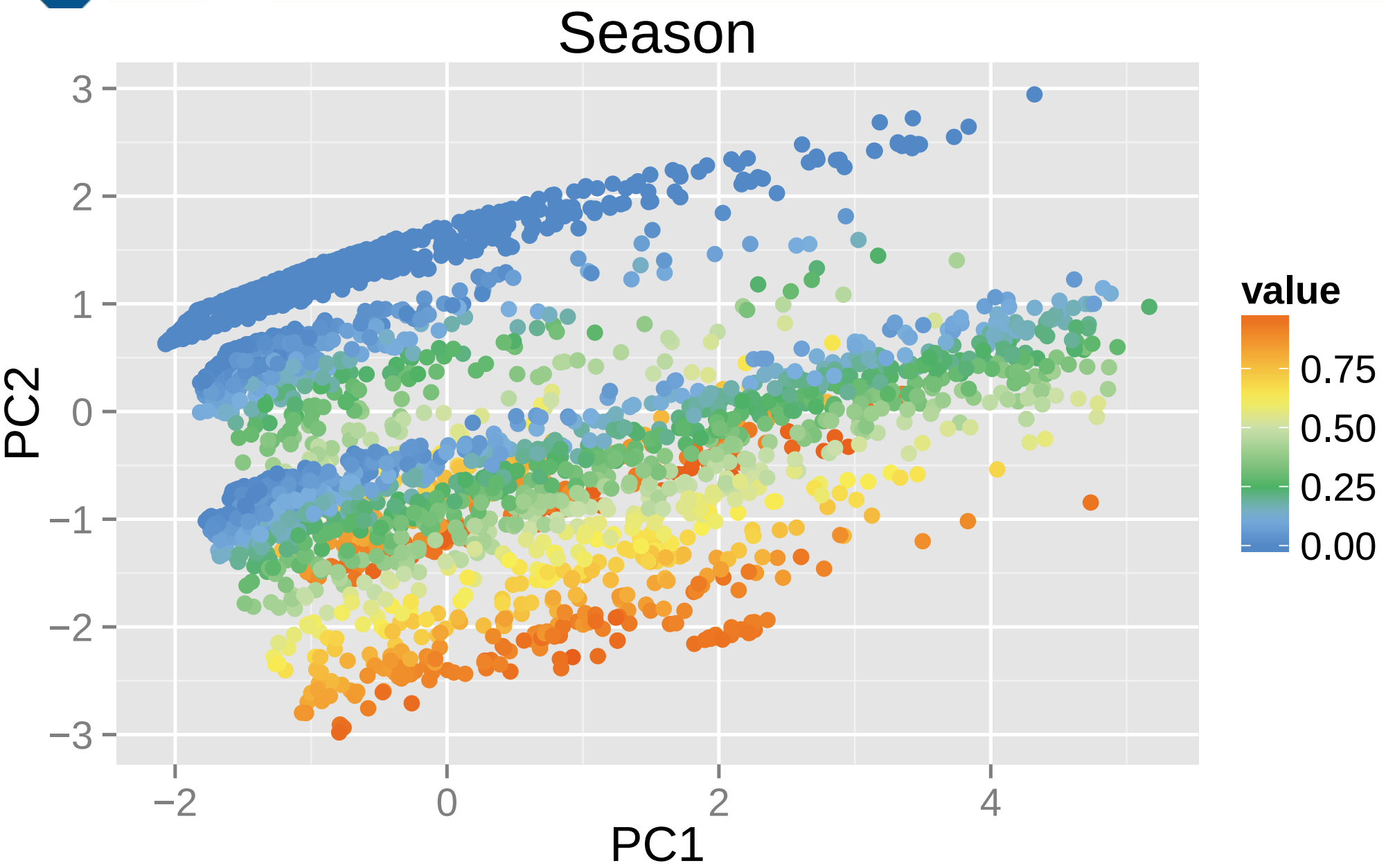
First two PCs explain 68% of variation.



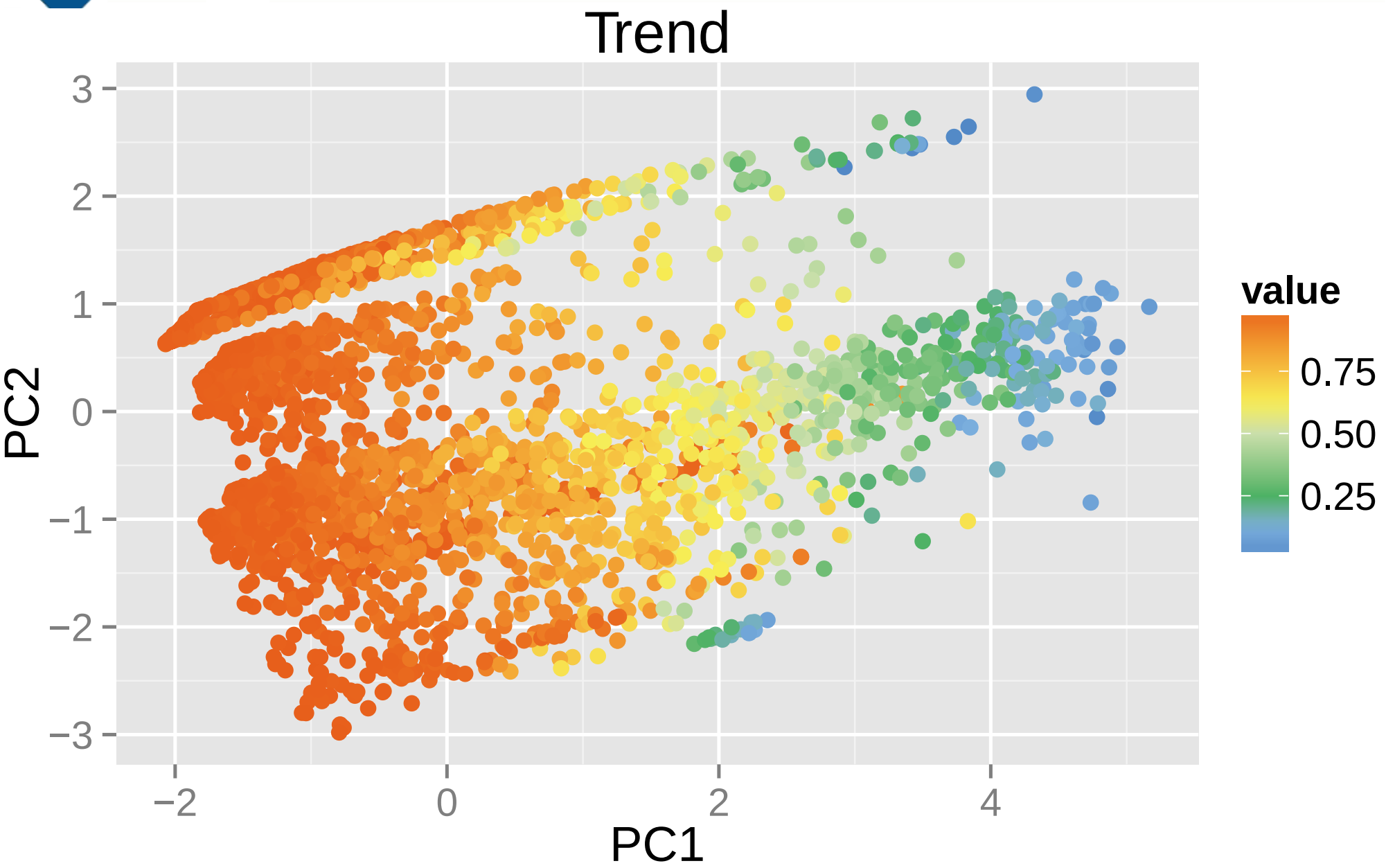
Feature space of M3 data



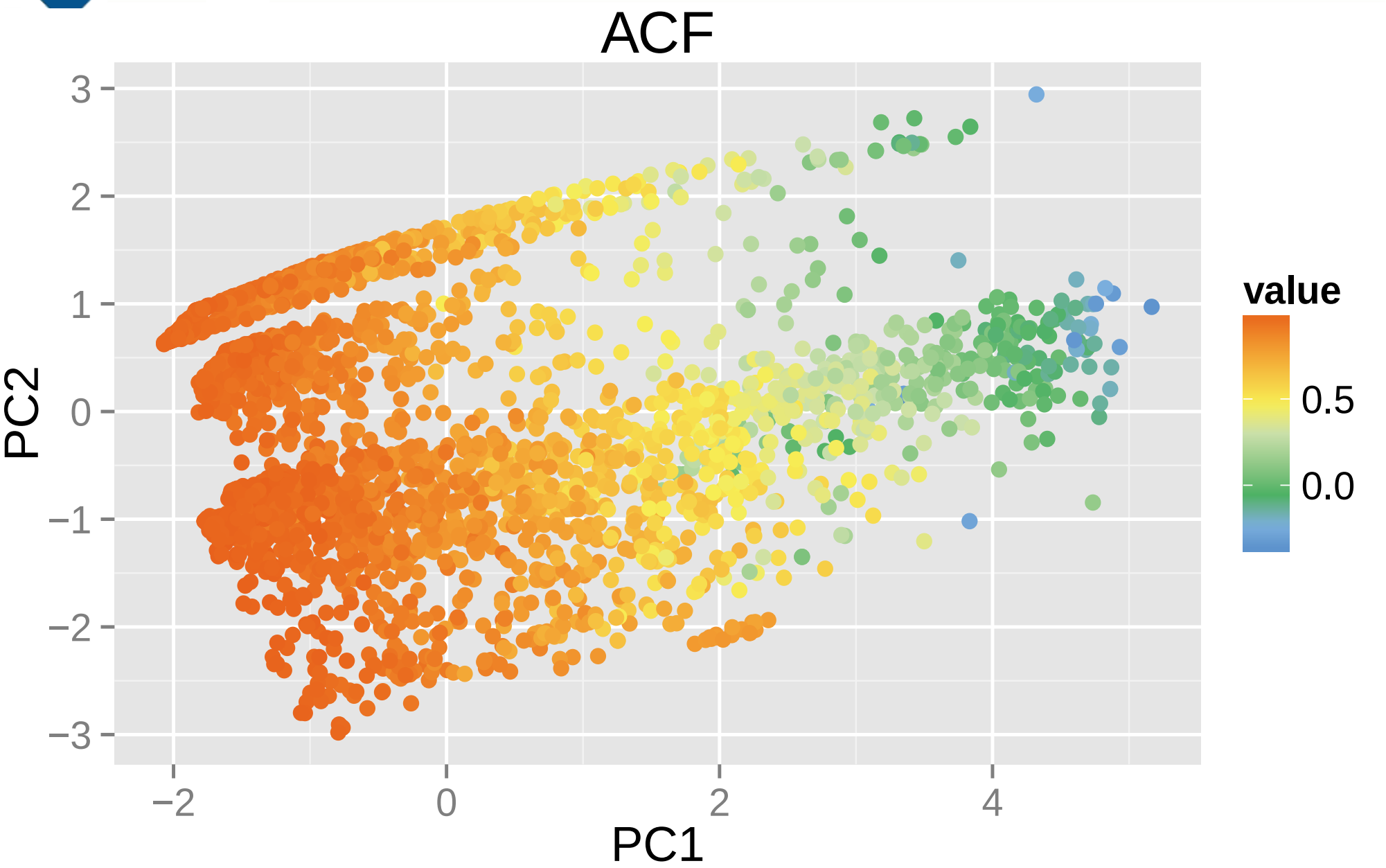
Feature space of M3 data



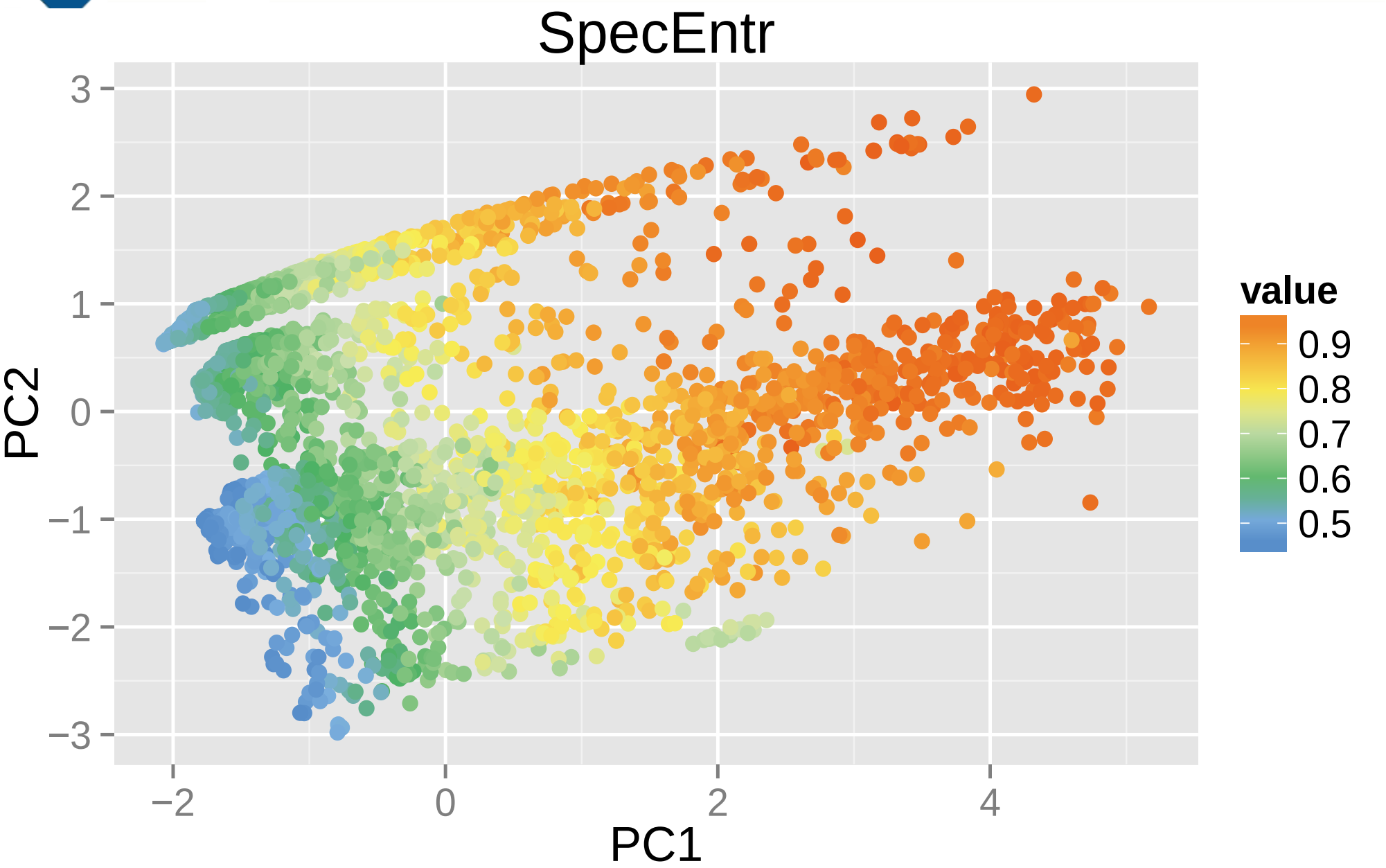
Feature space of M3 data



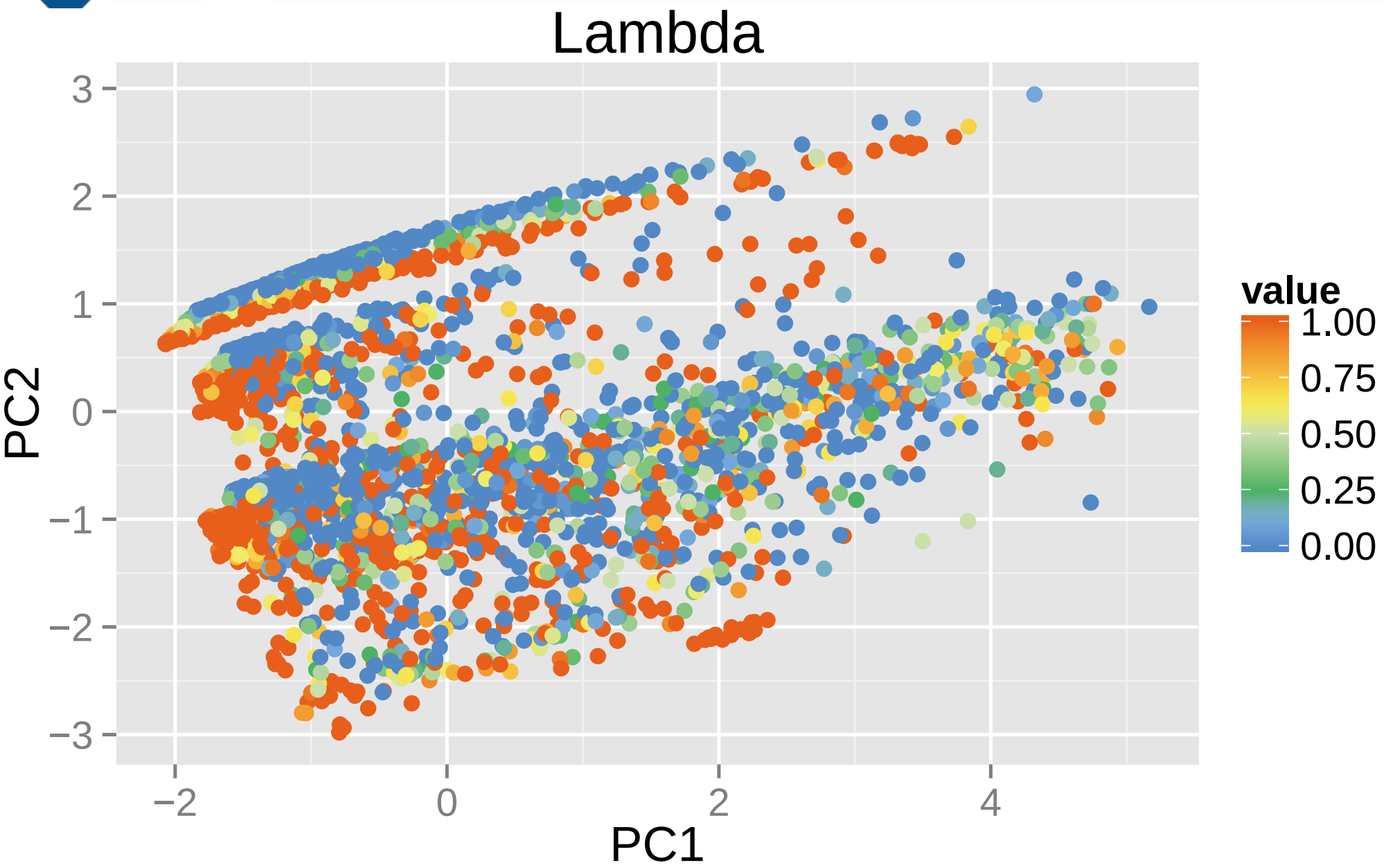
Feature space of M3 data



Feature space of M3 data



Feature space of M3 data

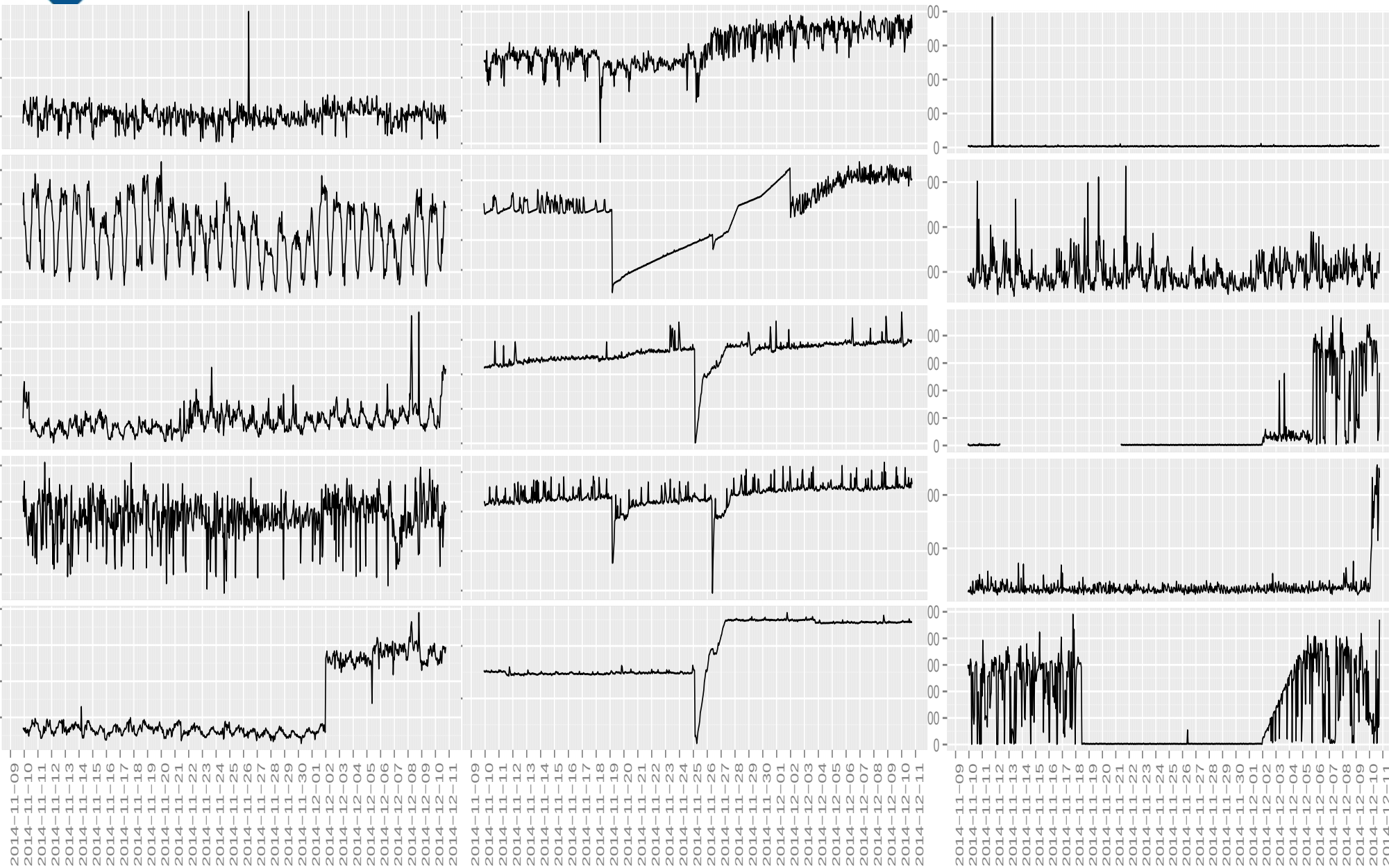


Outline

Yahoo web-traffic

- Tens of thousands of time series collected at one-hour intervals over one month.
- Consisting of several server metrics (e.g. CPU usage and paging views) from many server farms globally.
- Aim: find unusual (anomalous) time series.

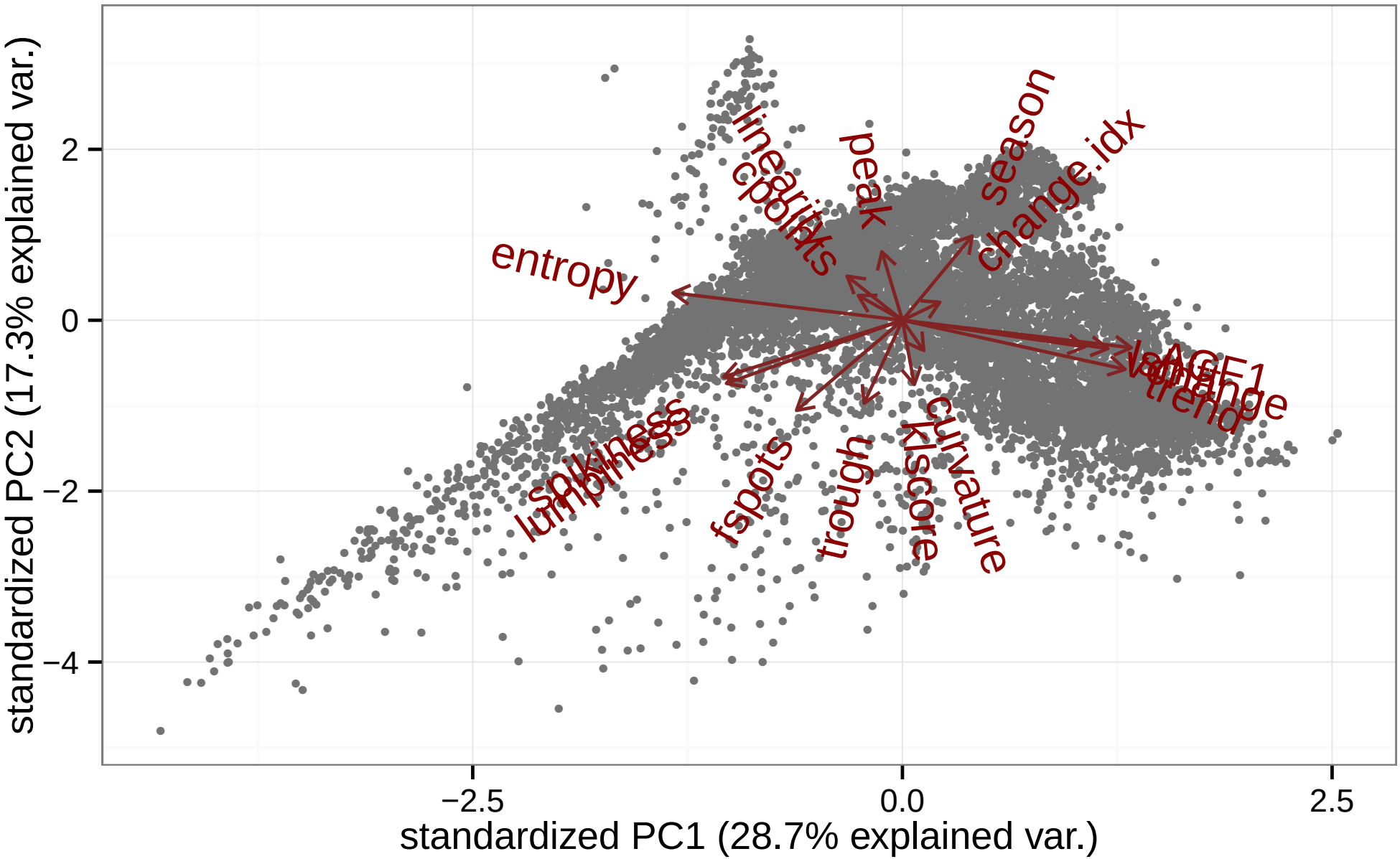
Yahoo web-traffic



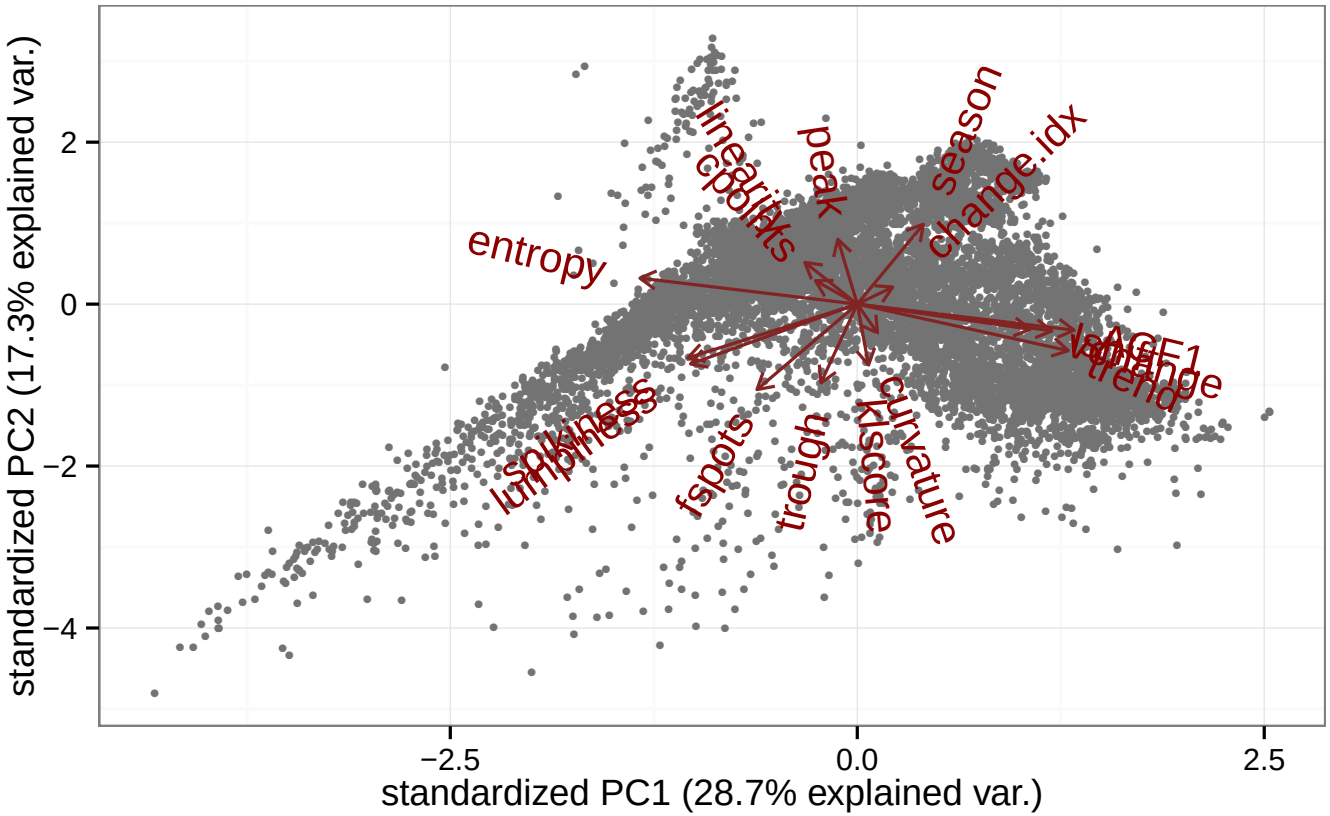
Feature space

- **ACF1**: first order autocorrelation = $\text{Corr}(Y_t, Y_{t-1})$
- Strength of **trend** and **seasonality** based on STL
- Trend **linearity** and **curvature**
- Size of seasonal **peak** and **trough**
- Spectral **entropy**
- **Lumpiness**: variance of block variances (block size 24).
- **Spikiness**: variances of leave-one-out variances of STL remainders.
- **Level shift**: Maximum difference in trimmed means of consecutive moving windows of size 24.
- **Variance change**: Max difference in variances of consecutive moving windows of size 24.
- **Flat spots**: Discretize sample space into 10 equal-sized intervals. Find max run length in any interval.
- Number of **crossing points** of mean line.
- **Kullback-Leibler score**: Maximum of $D_{KL}(P\|Q) = \int P(x) \ln P(x)/Q(x) dx$ where P and Q are estimated by kernel density estimators applied to consecutive windows of size 48.
- **Change index**: Time of maximum KL score

Principal component analysis



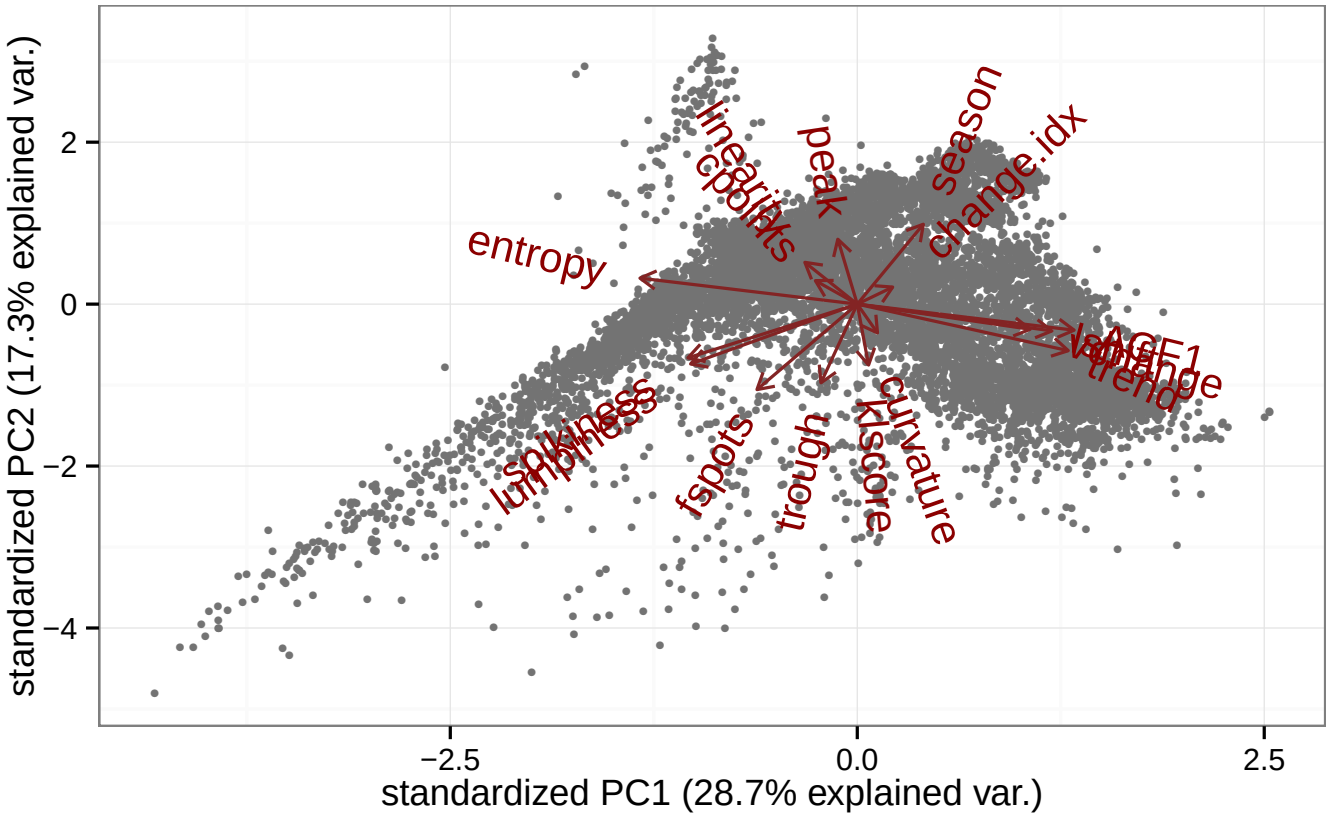
What is “anomalous”



We need a measure of the “anomalousness” of a time series.

1. Rank points based on their local density.
2. Rank points based on whether they are within α-convex hulls of different radius.

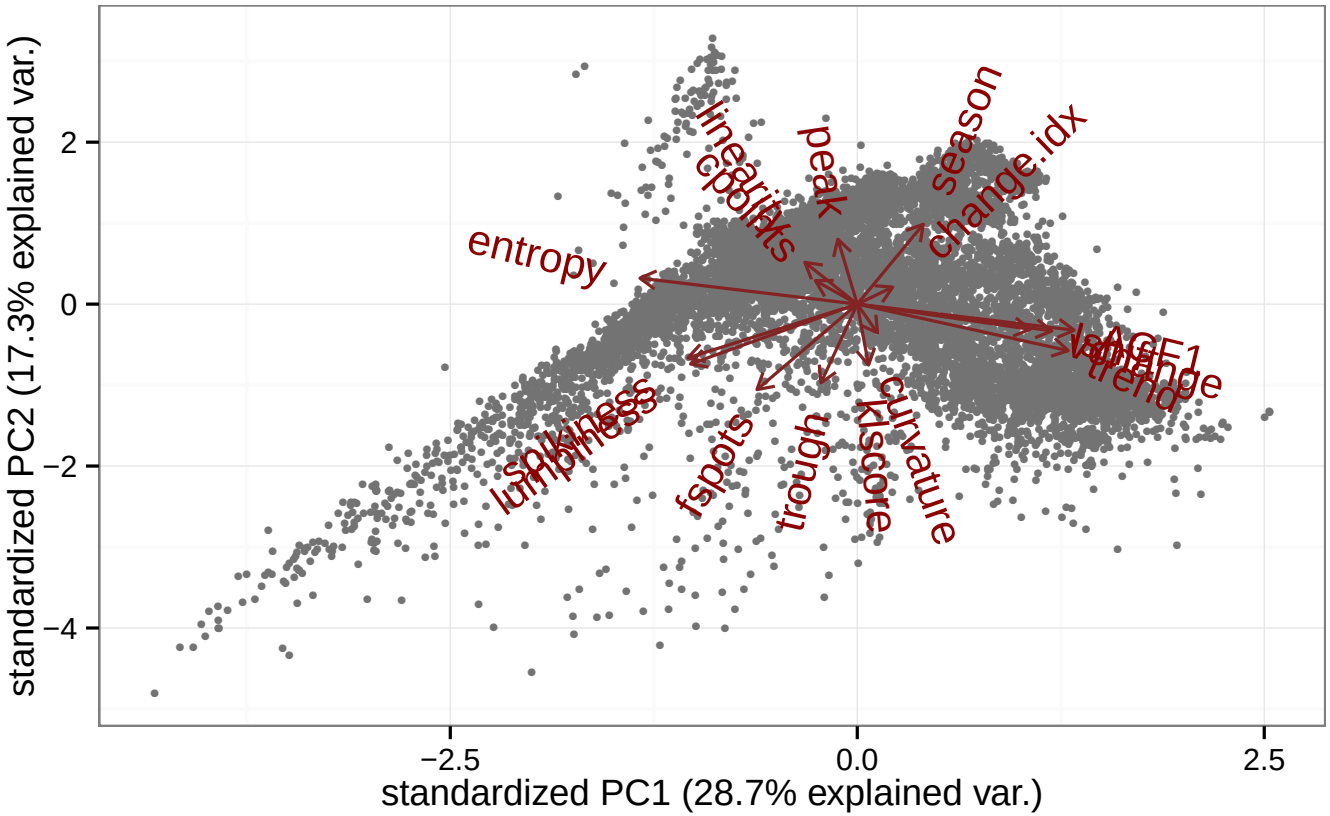
What is “anomalous”



We need a measure of the “anomalousness” of a time series.

1. Rank points based on their local density.
2. Rank points based on whether they are within α-convex hulls of different radius.

What is “anomalous”



We need a measure of the “anomalousness” of a time series.

1. Rank points based on their local density.
2. Rank points based on whether they are within α-convex hulls of different radius.

Bivariate kernel density

$$\hat{f}(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

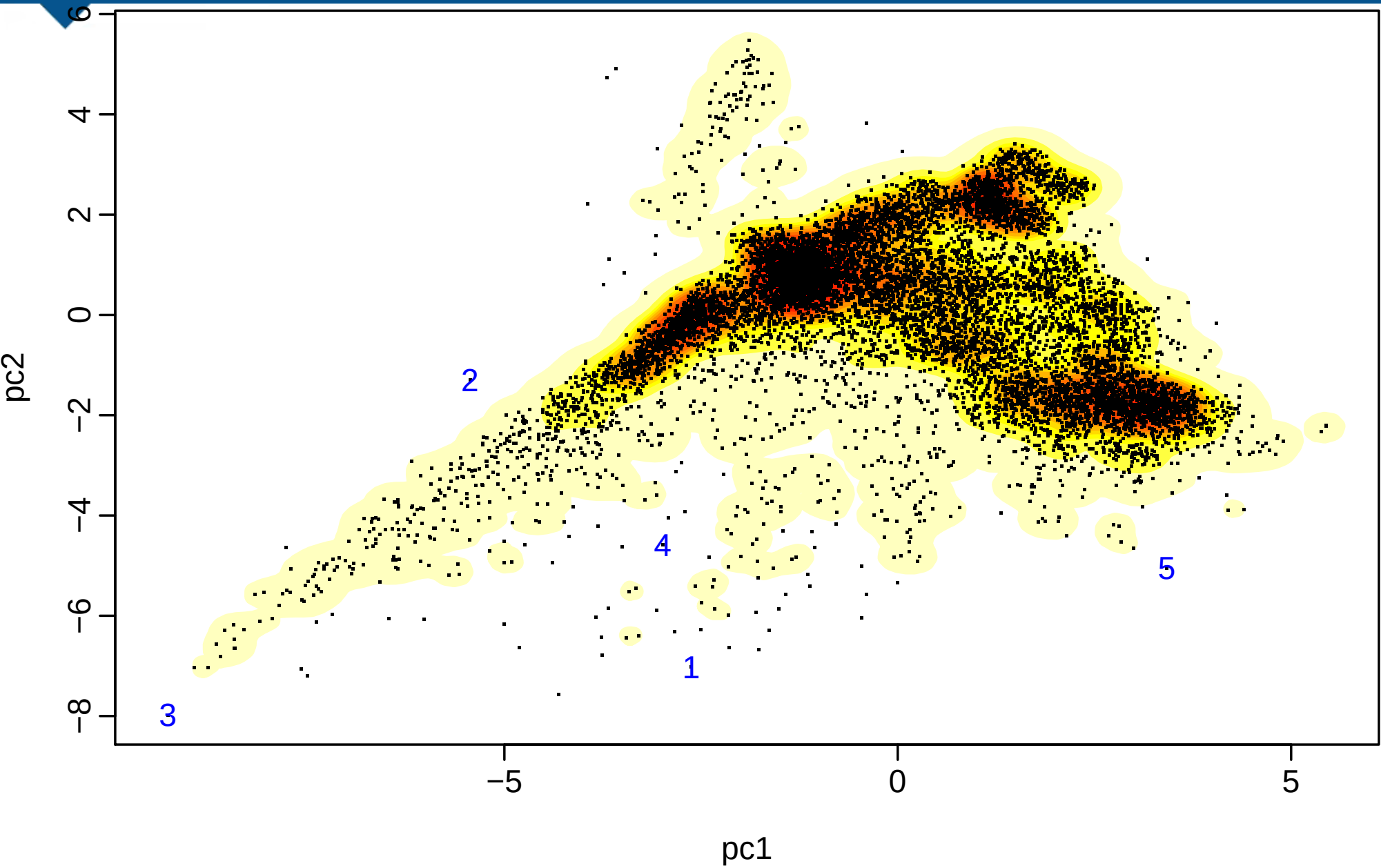
- $\mathbf{X}_i \in$ a bivariate random sample $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$
- $K_{\mathbf{H}}(\mathbf{x})$ is the standard normal kernel function
- $\mathbf{H}$ estimated by minimizing the sum of AMISE
- Rank points based on $\hat{f}$ values in 2d PCA space.

Bivariate kernel density

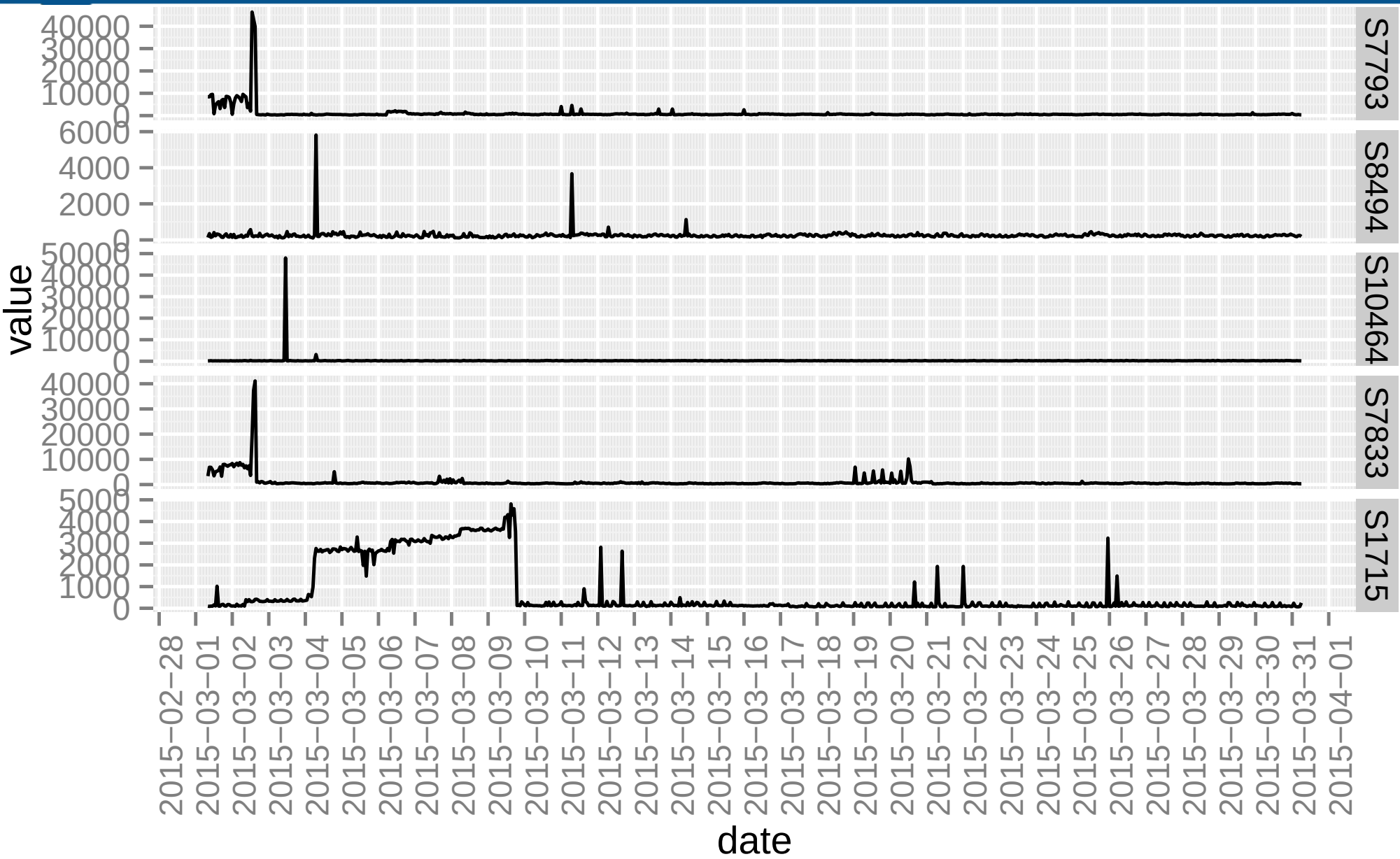
$$\hat{f}(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

- $\mathbf{X}_i \in$ a bivariate random sample $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$
- $K_{\mathbf{H}}(\mathbf{x})$ is the standard normal kernel function
- $\mathbf{H}$ estimated by minimizing the sum of AMISE
- Rank points based on $\hat{f}$ values in 2d PCA space.

Bivariate density ranking

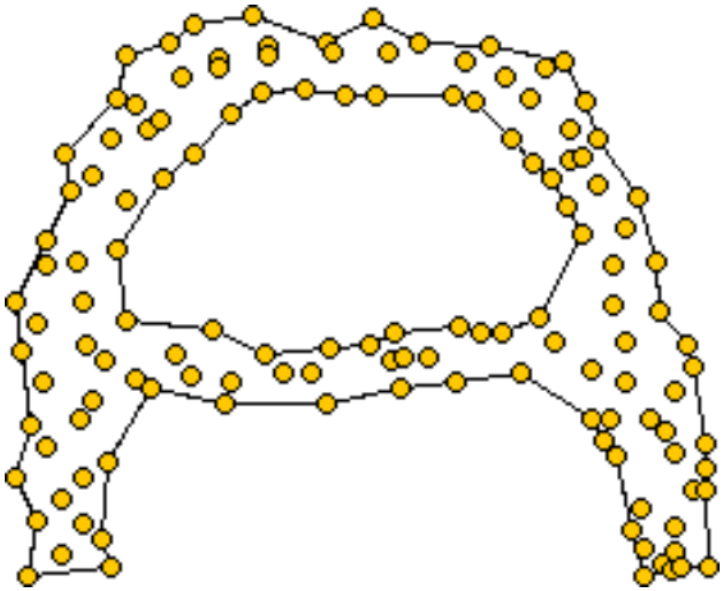


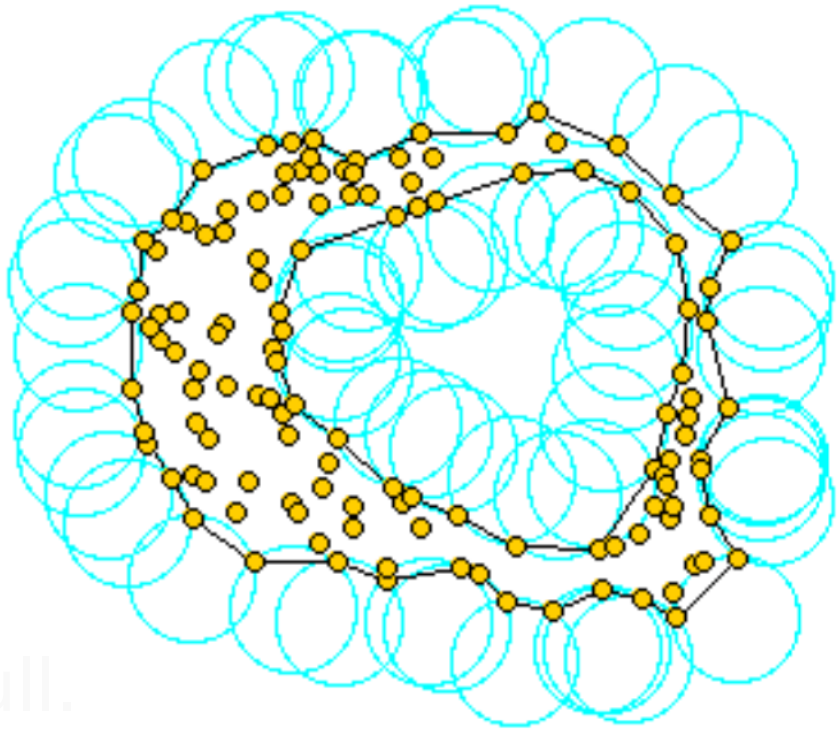
Bivariate density ranking



α-convex hulls

The space generated by point pairs that can be touched by an empty disc of radius α.

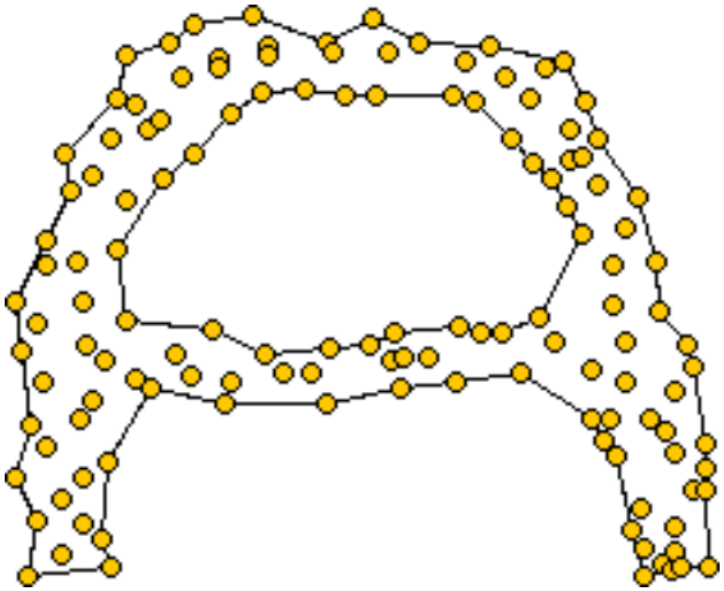


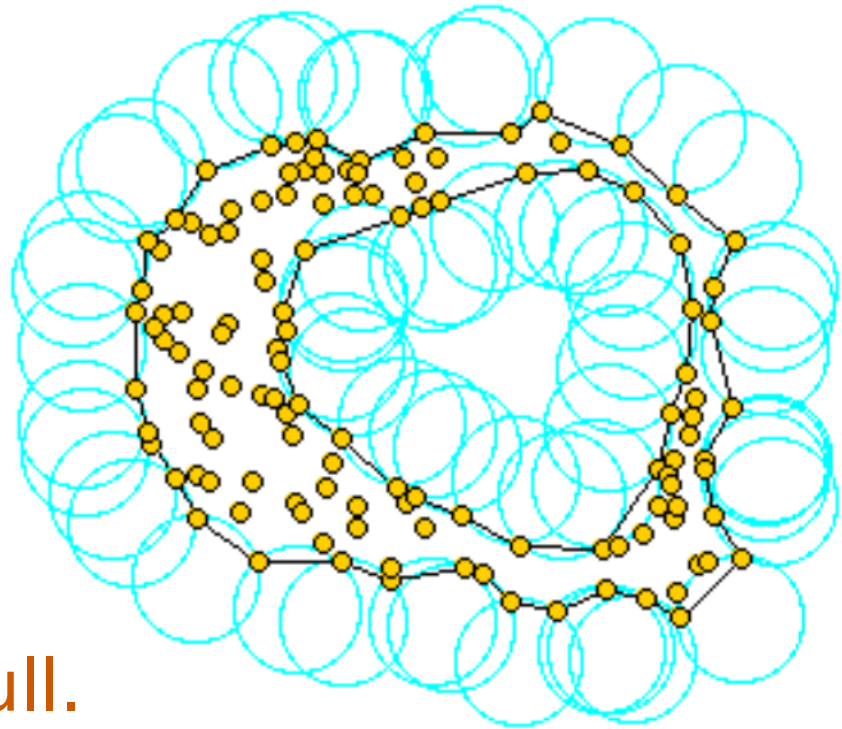


- $\alpha \to \infty$ gives a convex hull.
- Points can become isolated when α is small.
- We rank points based on the value of α when they become isolated.

α-convex hulls

The space generated by point pairs that can be touched by an empty disc of radius α.

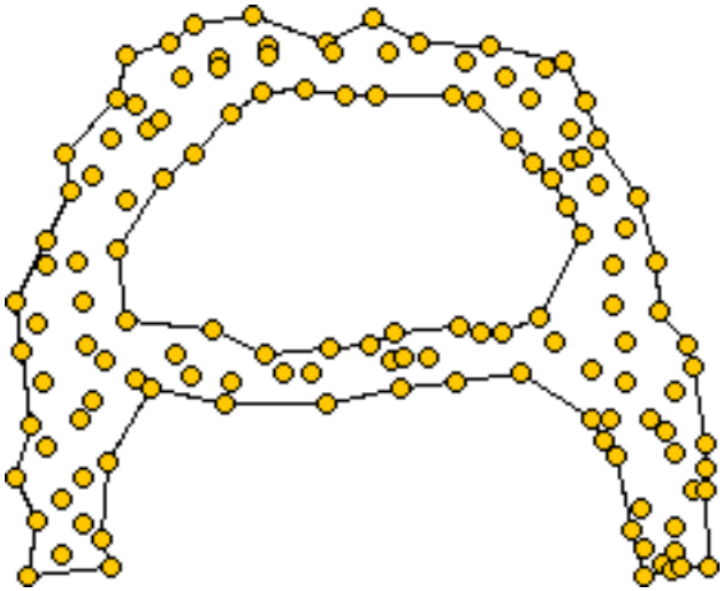


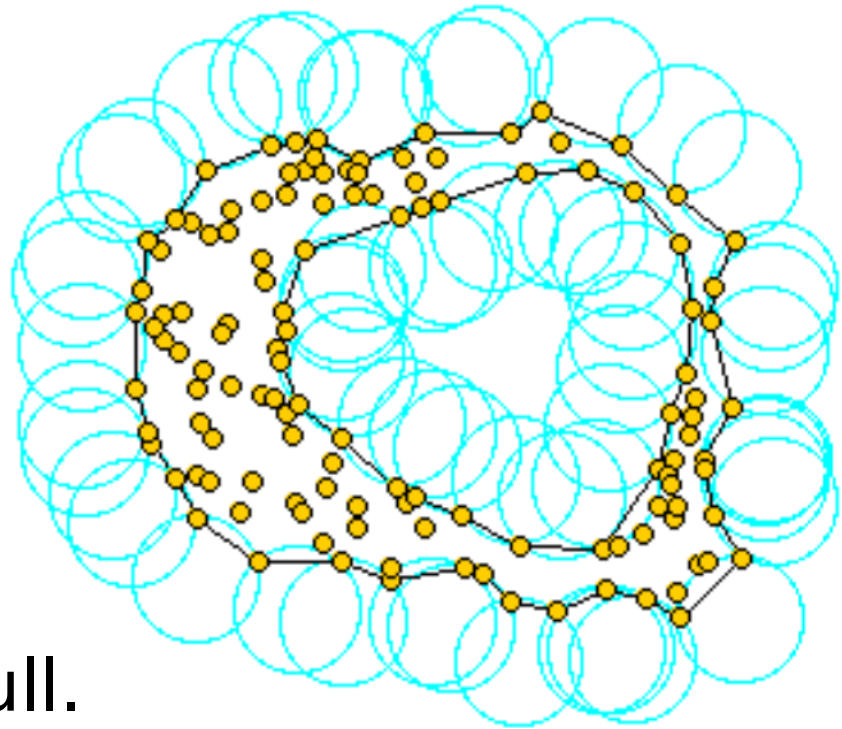


- $\alpha \to \infty$ gives a convex hull.
- Points can become isolated when α is small.
- We rank points based on the value of α when they become isolated.

α-convex hulls

The space generated by point pairs that can be touched by an empty disc of radius α.

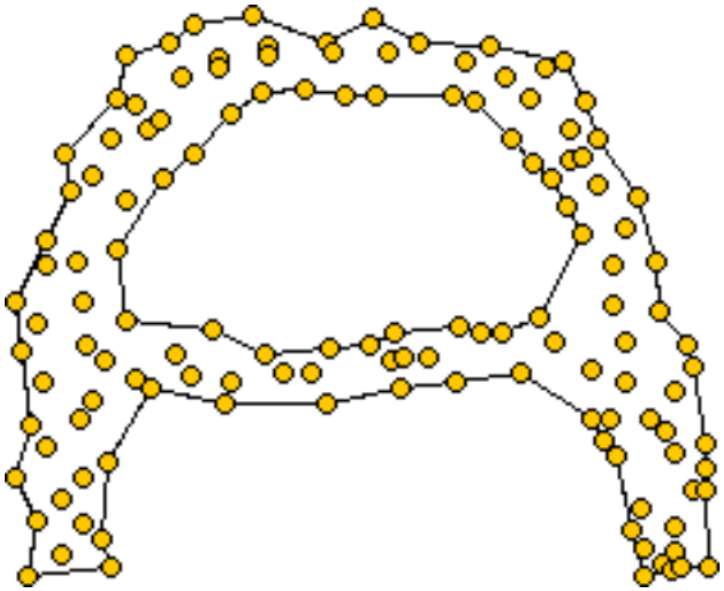


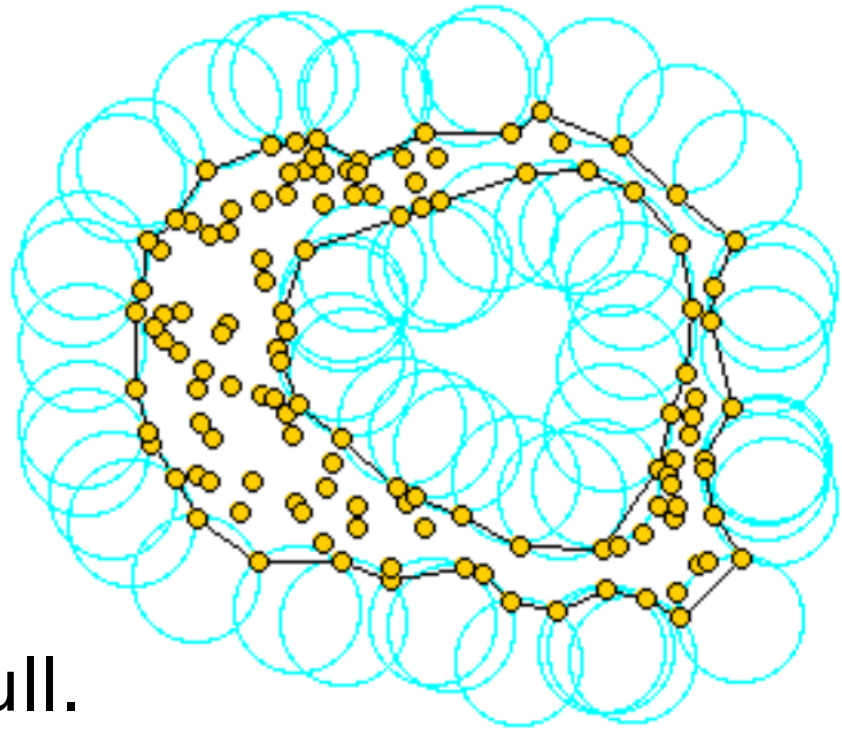


- $\alpha \to \infty$ gives a convex hull.
- Points can become isolated when α is small.
- We rank points based on the value of α when they become isolated.

α-convex hulls

The space generated by point pairs that can be touched by an empty disc of radius α.

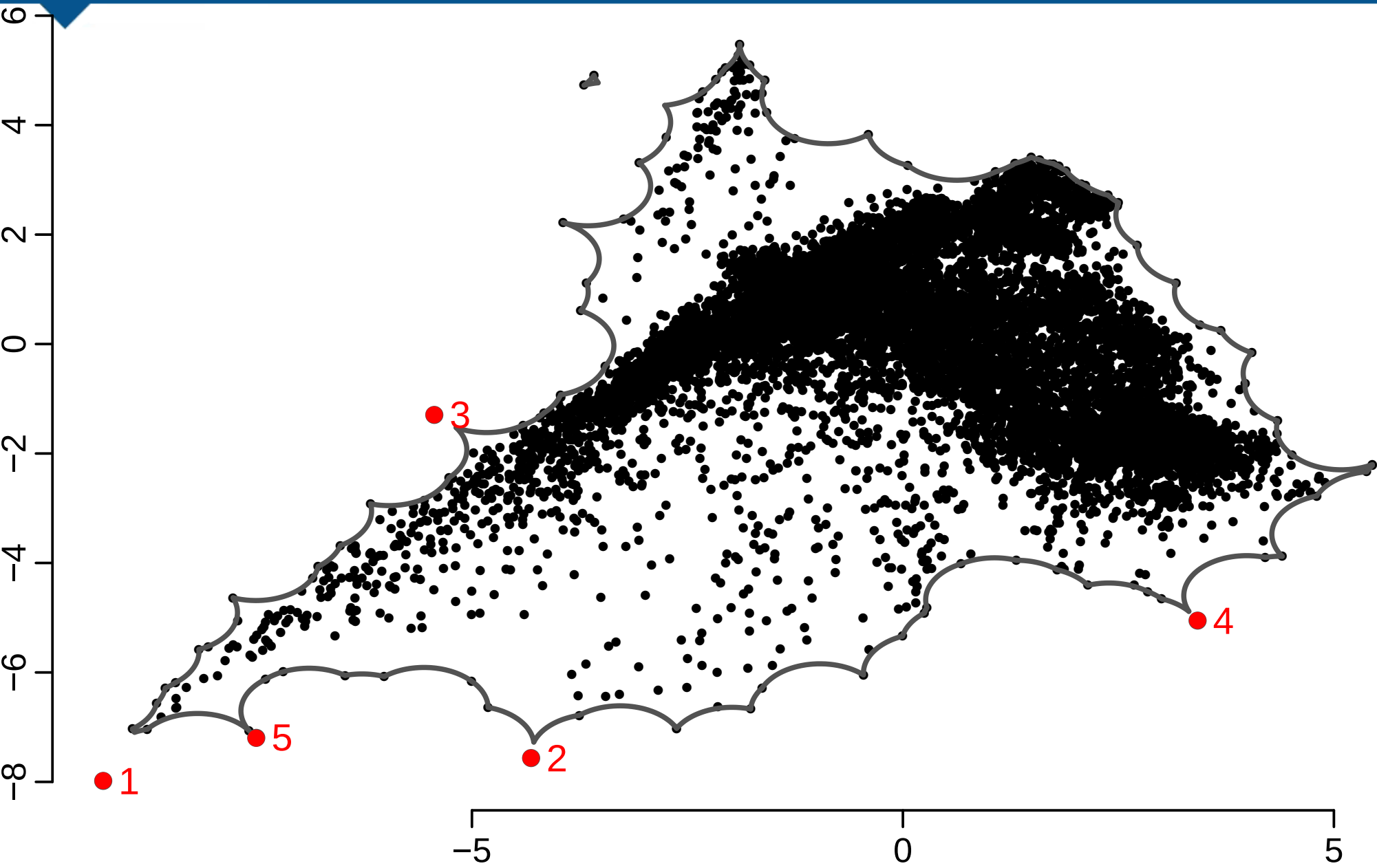




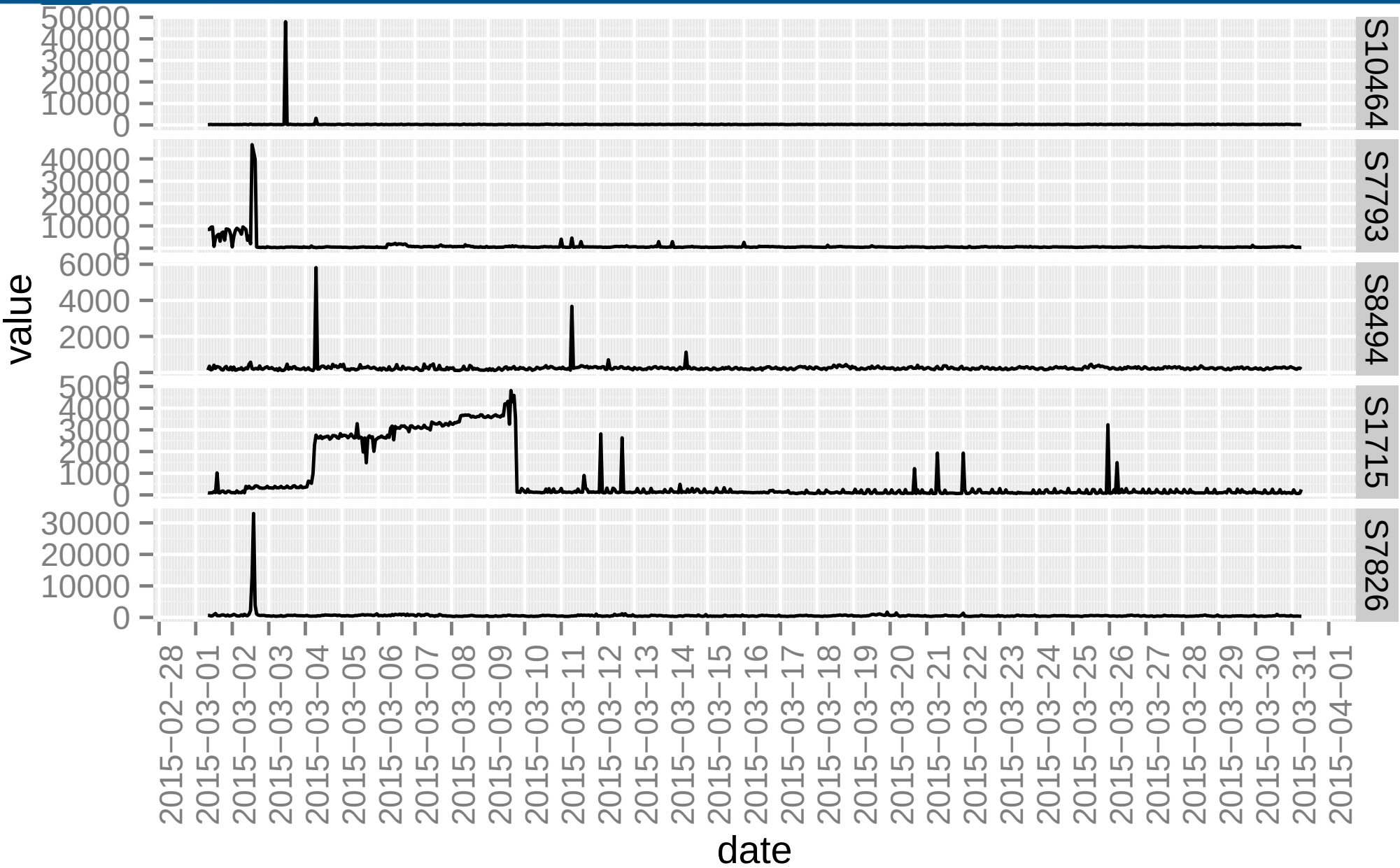
- $\alpha \to \infty$ gives a convex hull.
- Points can become isolated when α is small.
- We rank points based on the value of α when they become isolated.

α-convex hull

α-convex hull ranking

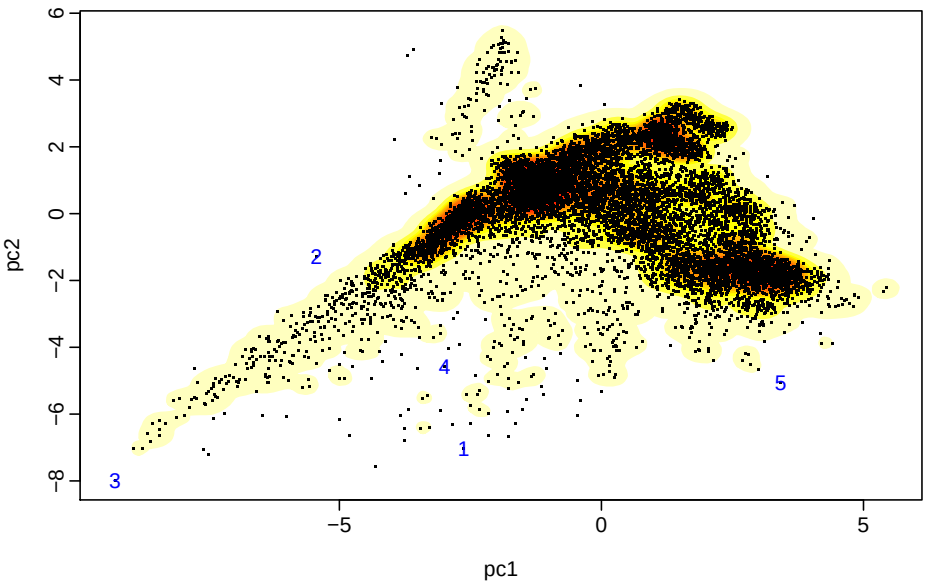


α-convex hull ranking

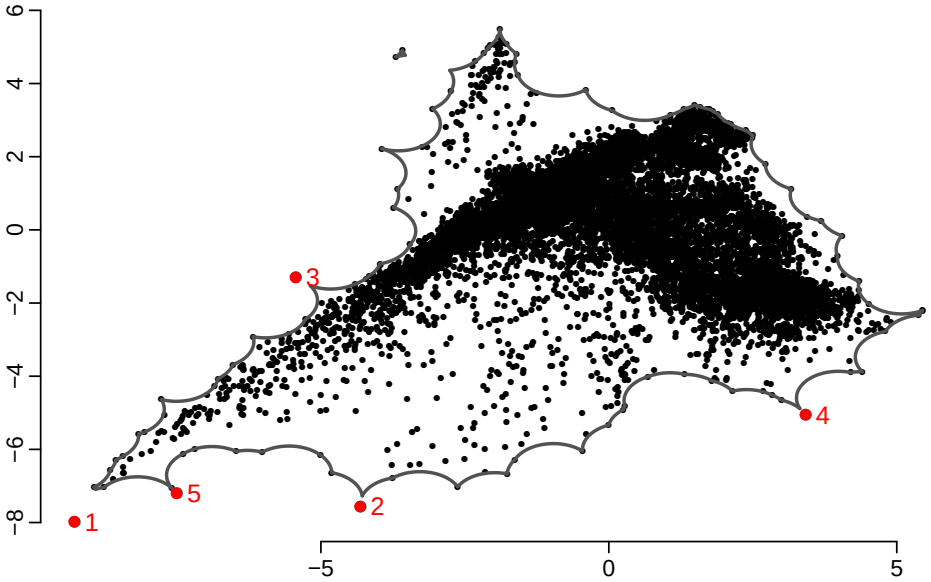


HDR versus α-convex hull

HDR boxplot

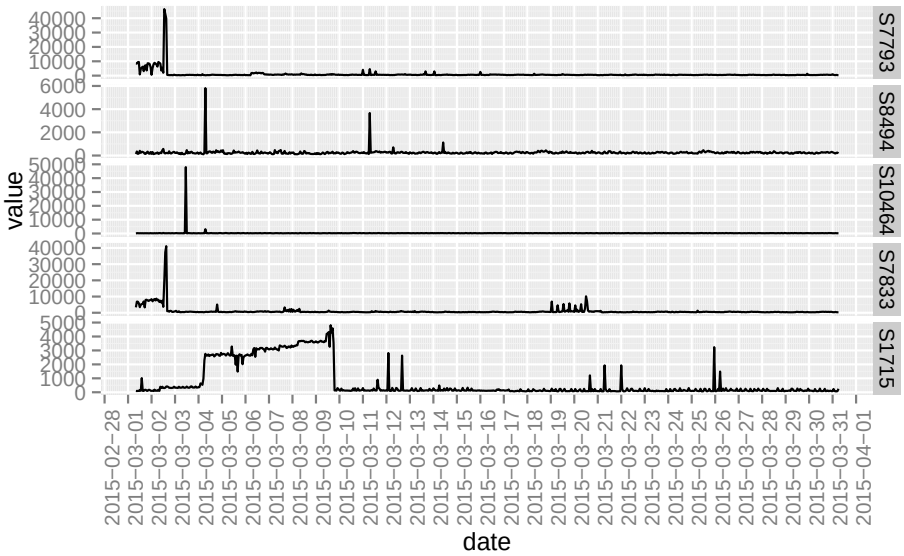


α-convex hull

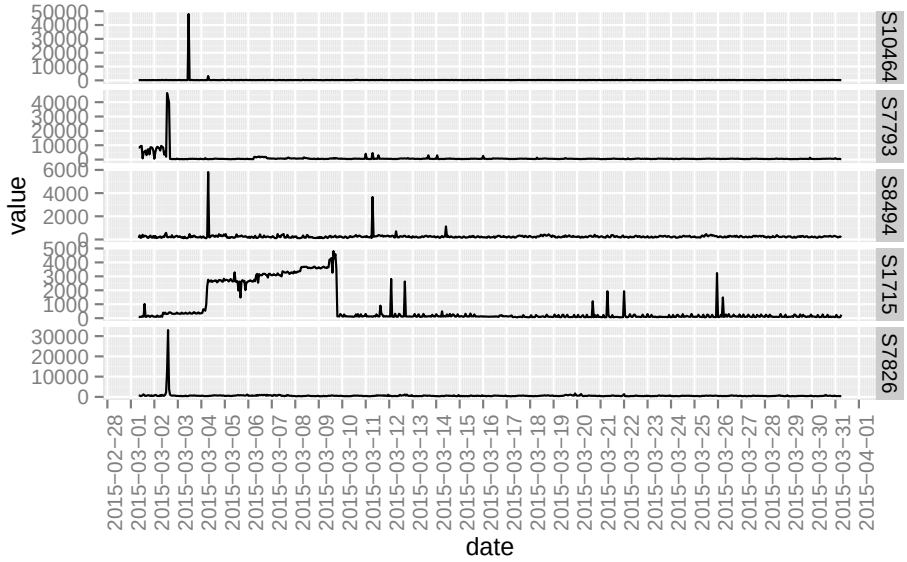


Top 5 anomalous time series

HDR



α-convex hull



Outline

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.

Some of the methods are also available in the anomalous package for R on github.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.
  - Some of the methods are already available in the **anomalous** package for R on github.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.
  - Some of the methods are already available in the **anomalous** package for R on github.

What next?

- Develop a more comprehensive set of features that are reliable measures and fast to compute. e.g., for finance data.
- Consider application to functional data from other contexts (not time series).
- Is PCA the right approach? Perhaps we should use multidimensional scaling? Or something else?
- Should we use more than 2 PC dimensions?
- Develop dynamic and interactive visualization tools.
- Make methods available in an R package.
  - Some of the methods are already available in the **anomalous** package for R on github.

Further information

- Papers: **robjhyndman.com**
- Code: **github.com/robjhyndman**
- Email: **Rob.Hyndman@monash.edu**