

Feature-based time series analysis

Rob J Hyndman

27 September 2019

Outline

- 1 Feature-based visualization
- 2 R packages
- 3 Feature-based anomaly detection
- 4 Feature-based forecasting

Outline

- 1 Feature-based visualization
- 2 R packages
- 3 Feature-based anomaly detection
- 4 Feature-based forecasting

M3 competition: 2000



ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

www.elsevier.com/locate/ijforecast

*international journal
of forecasting*

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. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting 4 accuracy

M3 competition: 2000



ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

*international journal
of forecasting*

www.elsevier.com/locate/ijforecast



petition: results, conclusions a

Spyros Makridakis, Michèle Hibon*

EAD, Boulevard de Constance, 77305 Fontainebleau, Fr

Abst



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. © 2000 Elsevier Science B.V. All rights reserved.

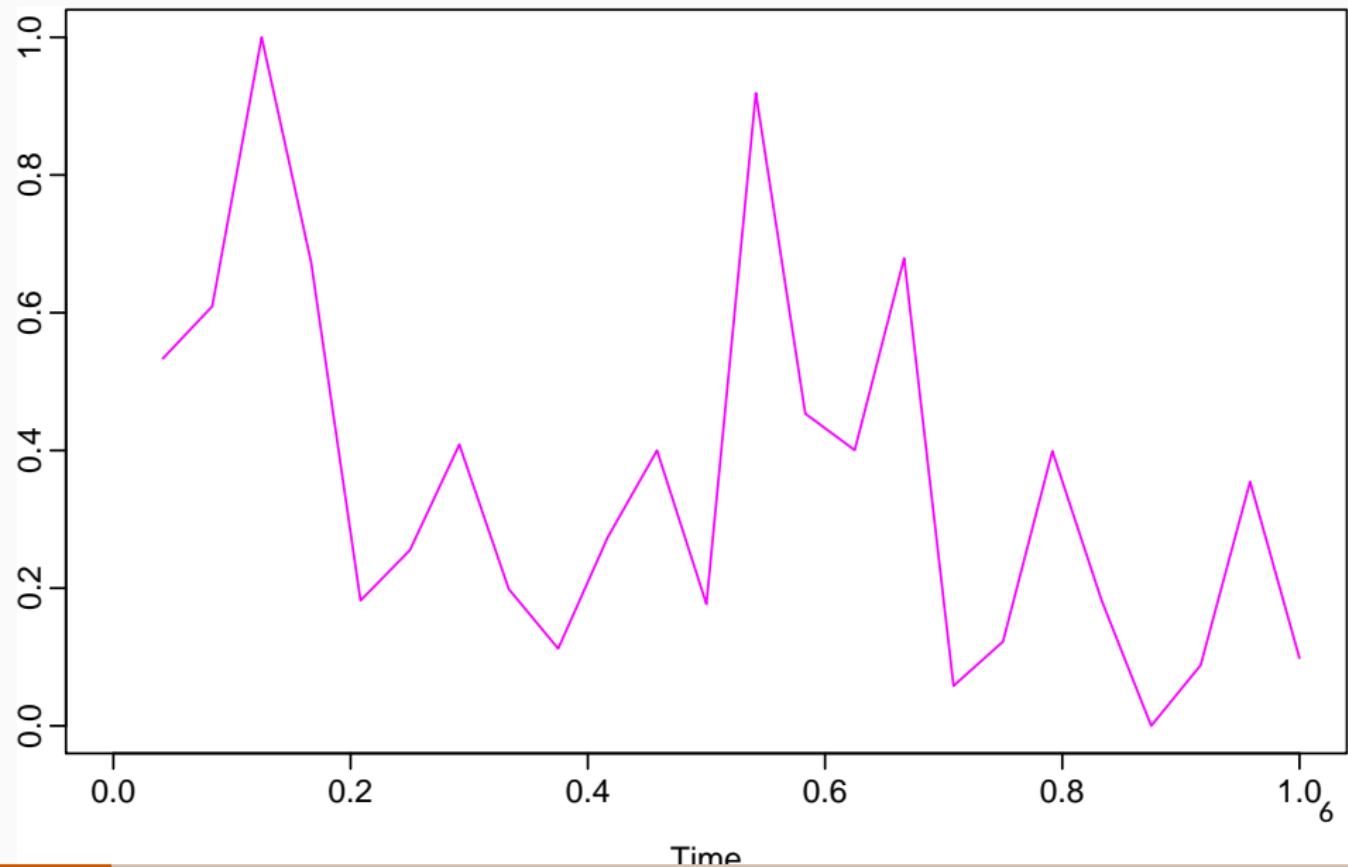
Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting accuracy 4

M3 competition: 2000

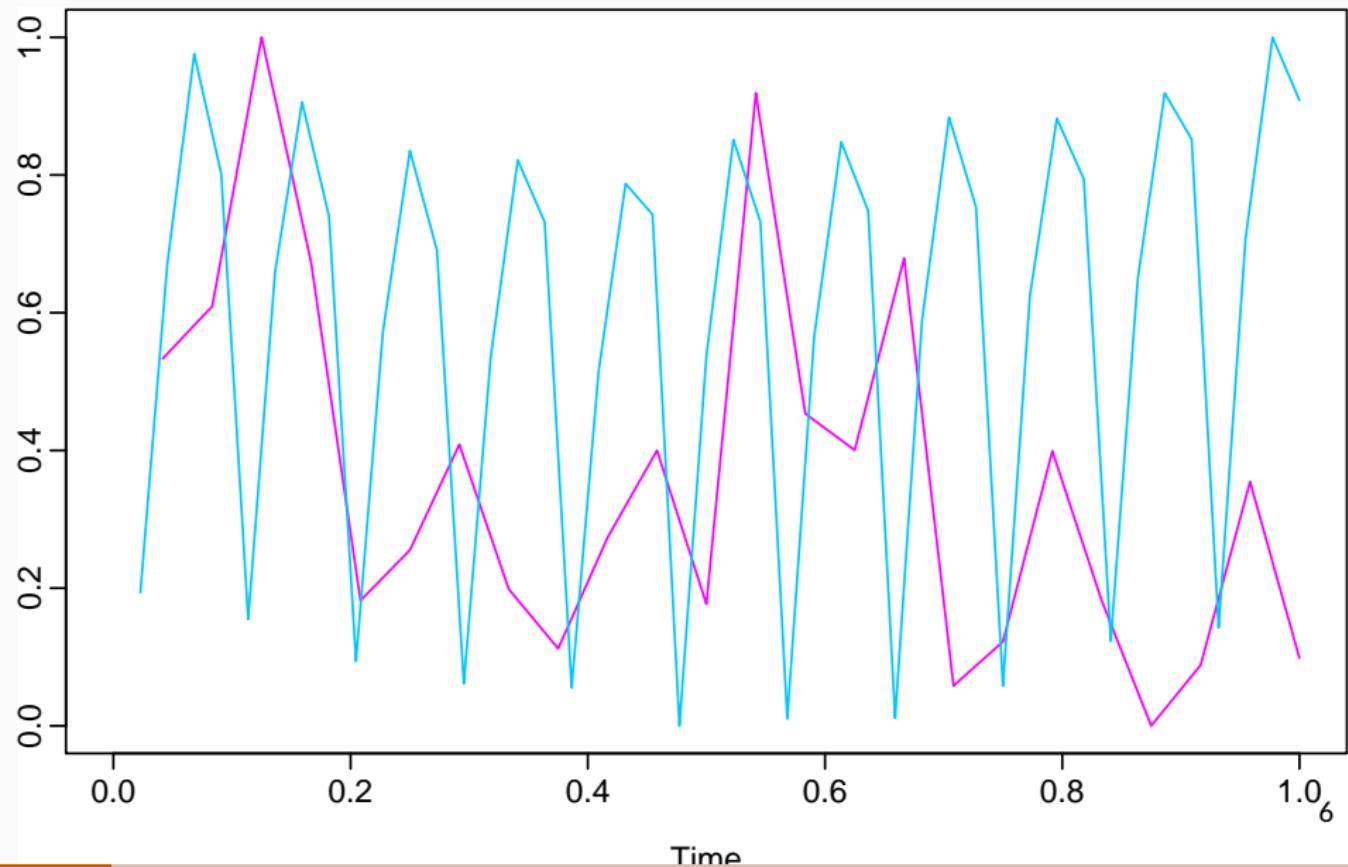
“The M3-Competition is a final attempt by the authors to settle the accuracy issue of various time series methods... The extension involves the inclusion of more methods/researchers (in particular in the areas of neural networks and expert systems) and more series.”

- 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.

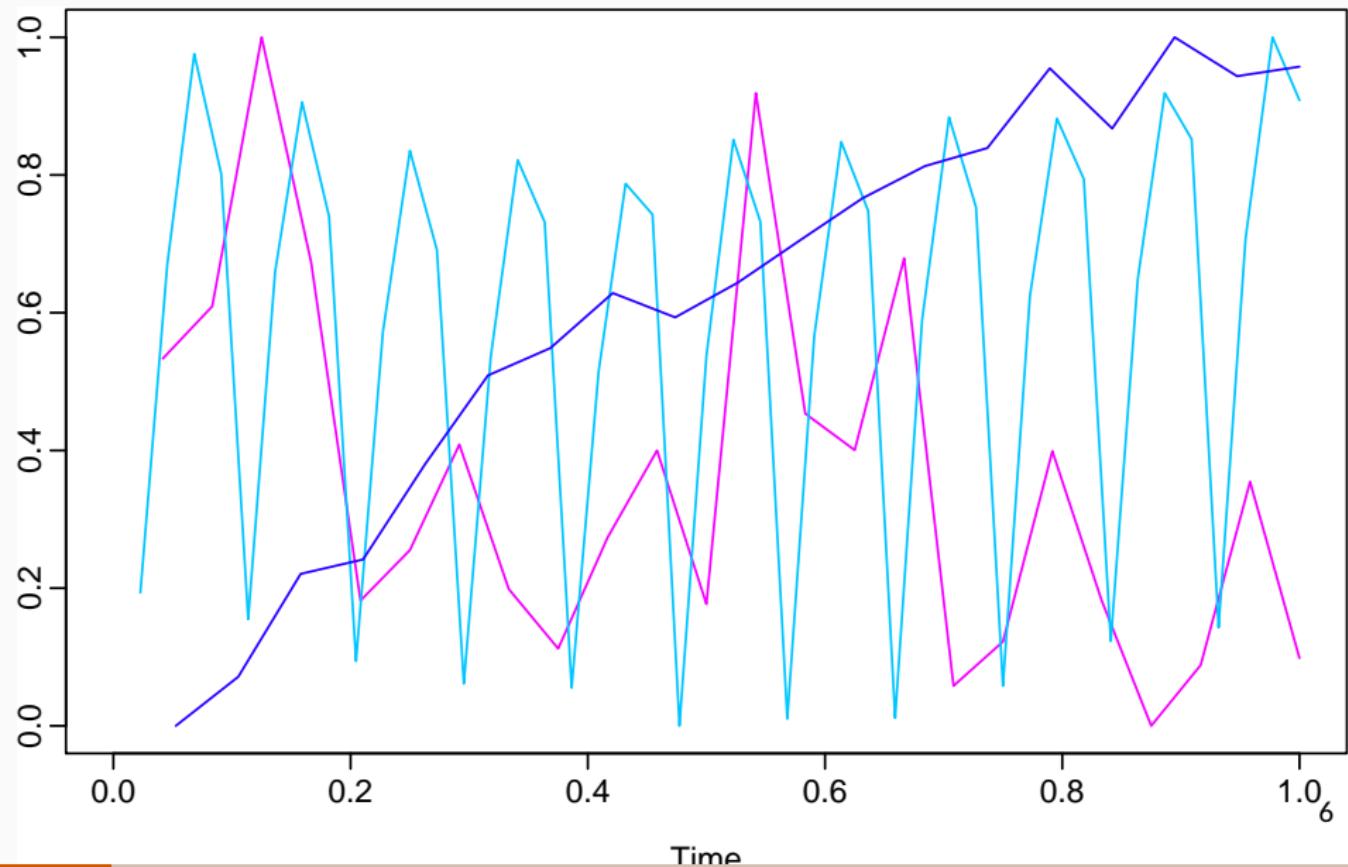
How to plot lots of time series?



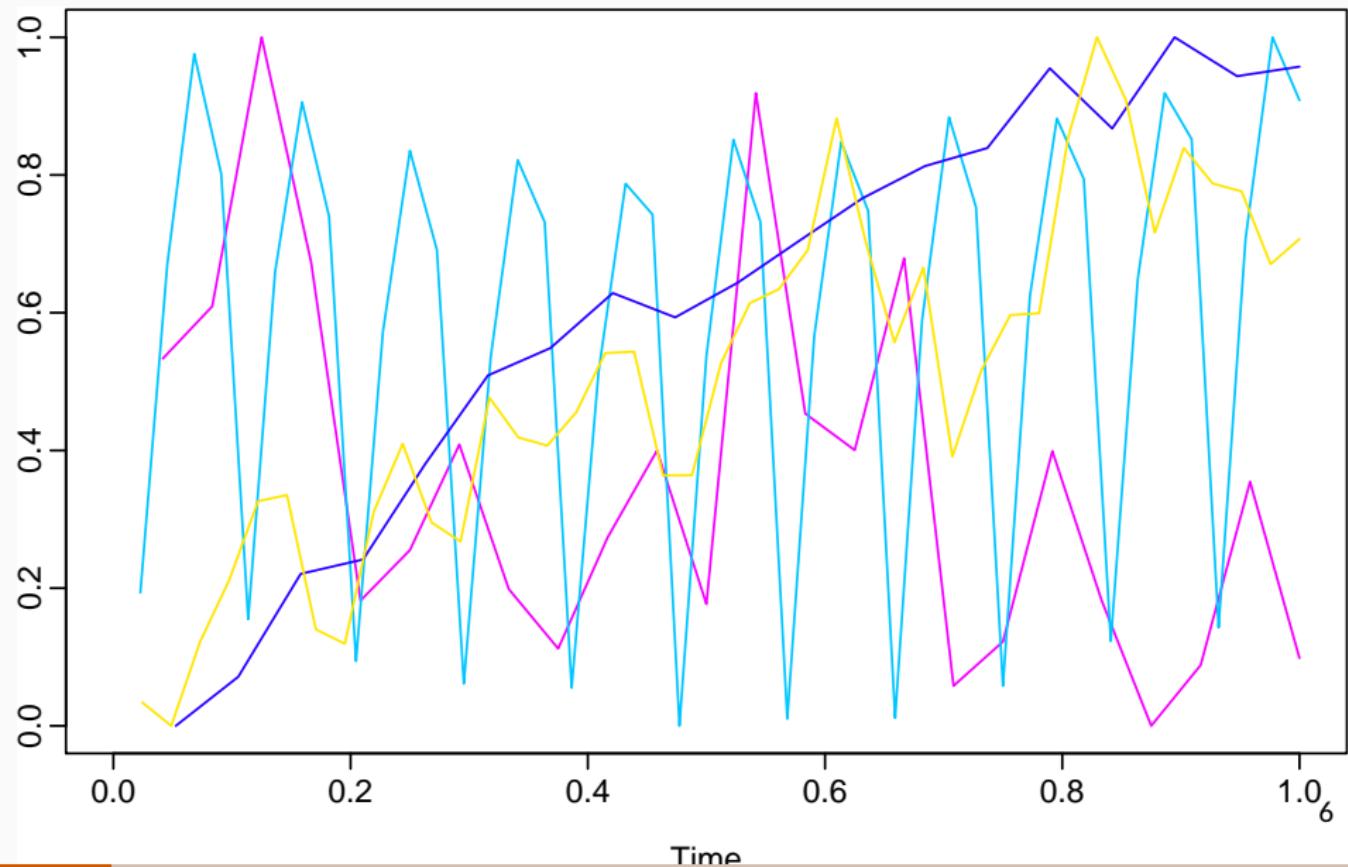
How to plot lots of time series?



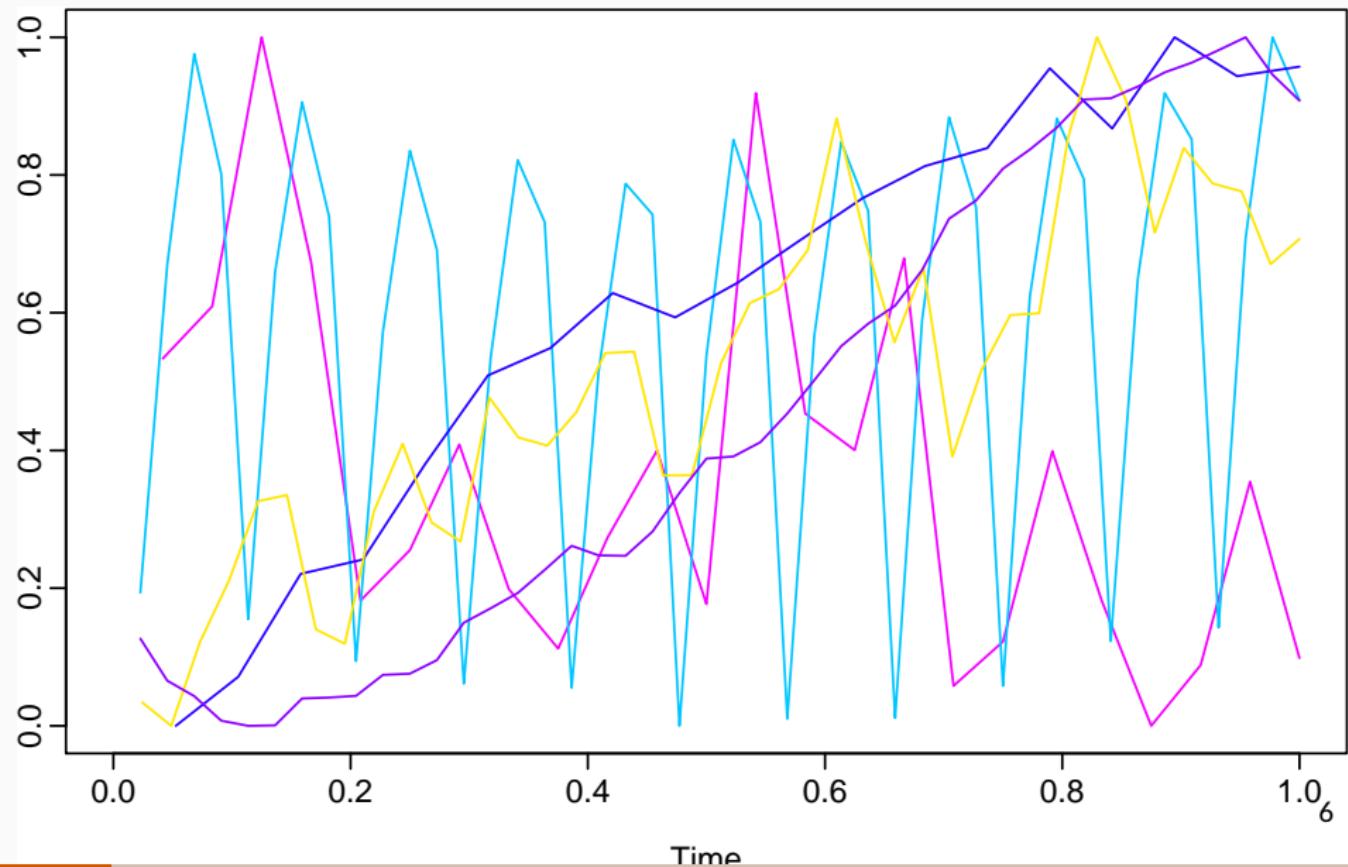
How to plot lots of time series?



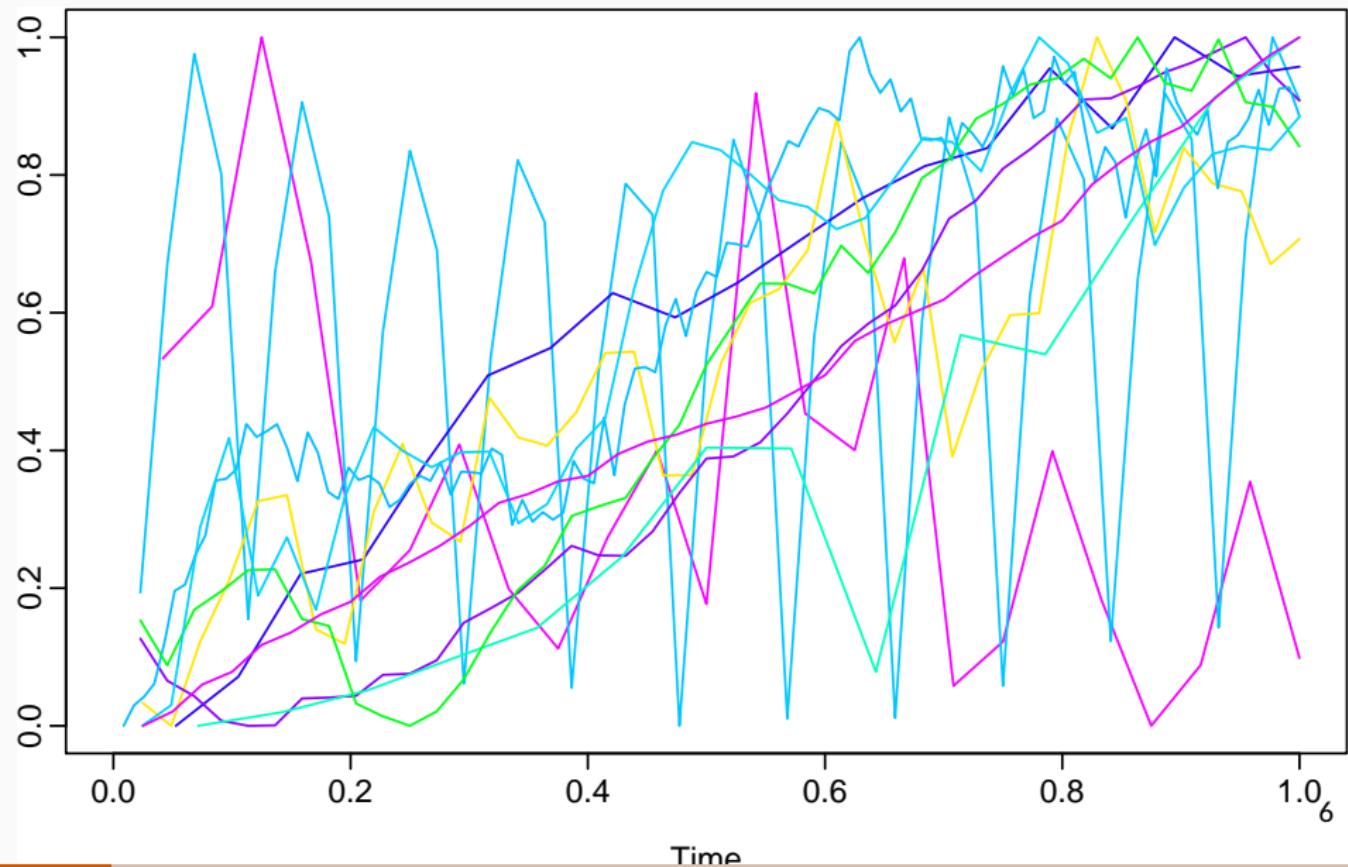
How to plot lots of time series?



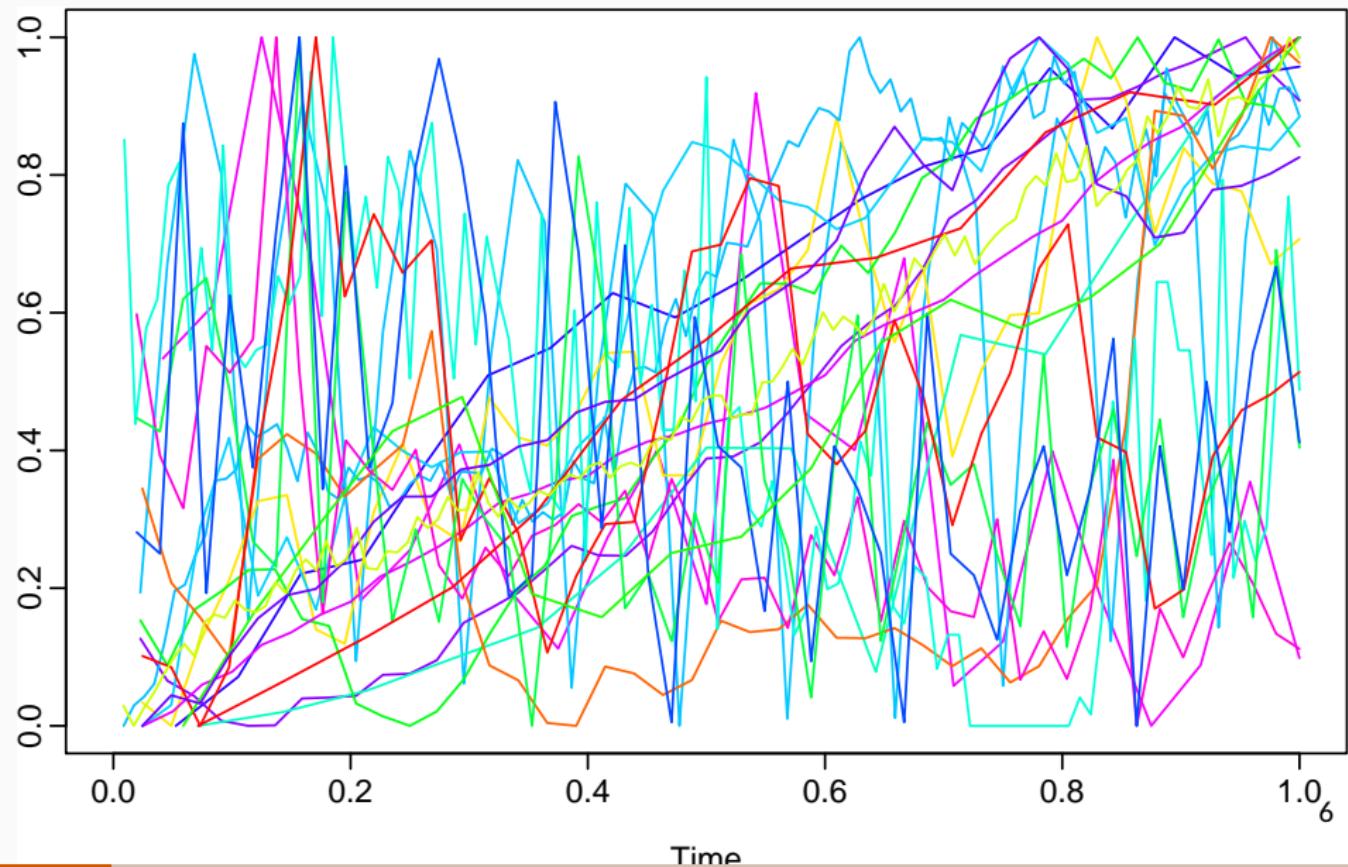
How to plot lots of time series?



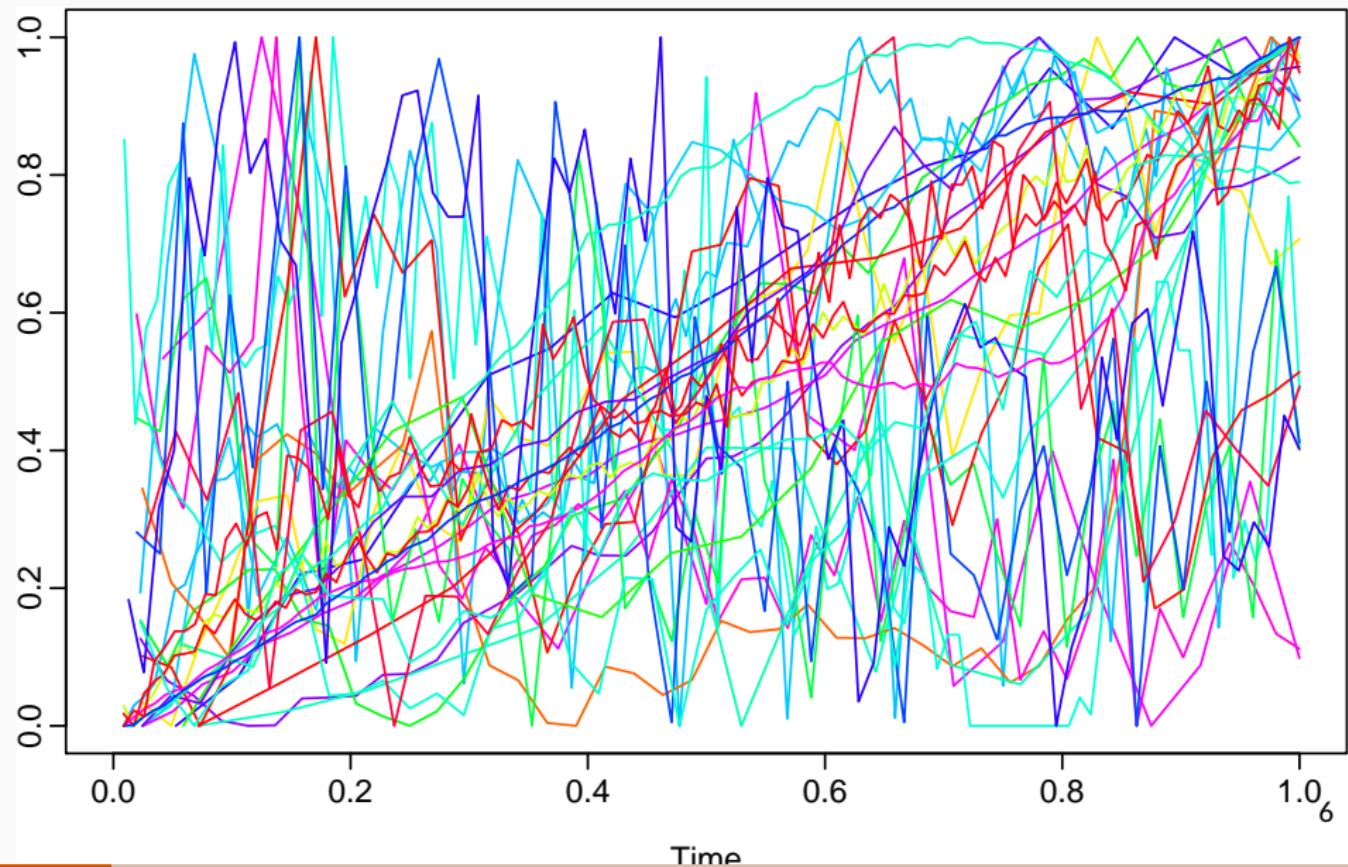
How to plot lots of time series?



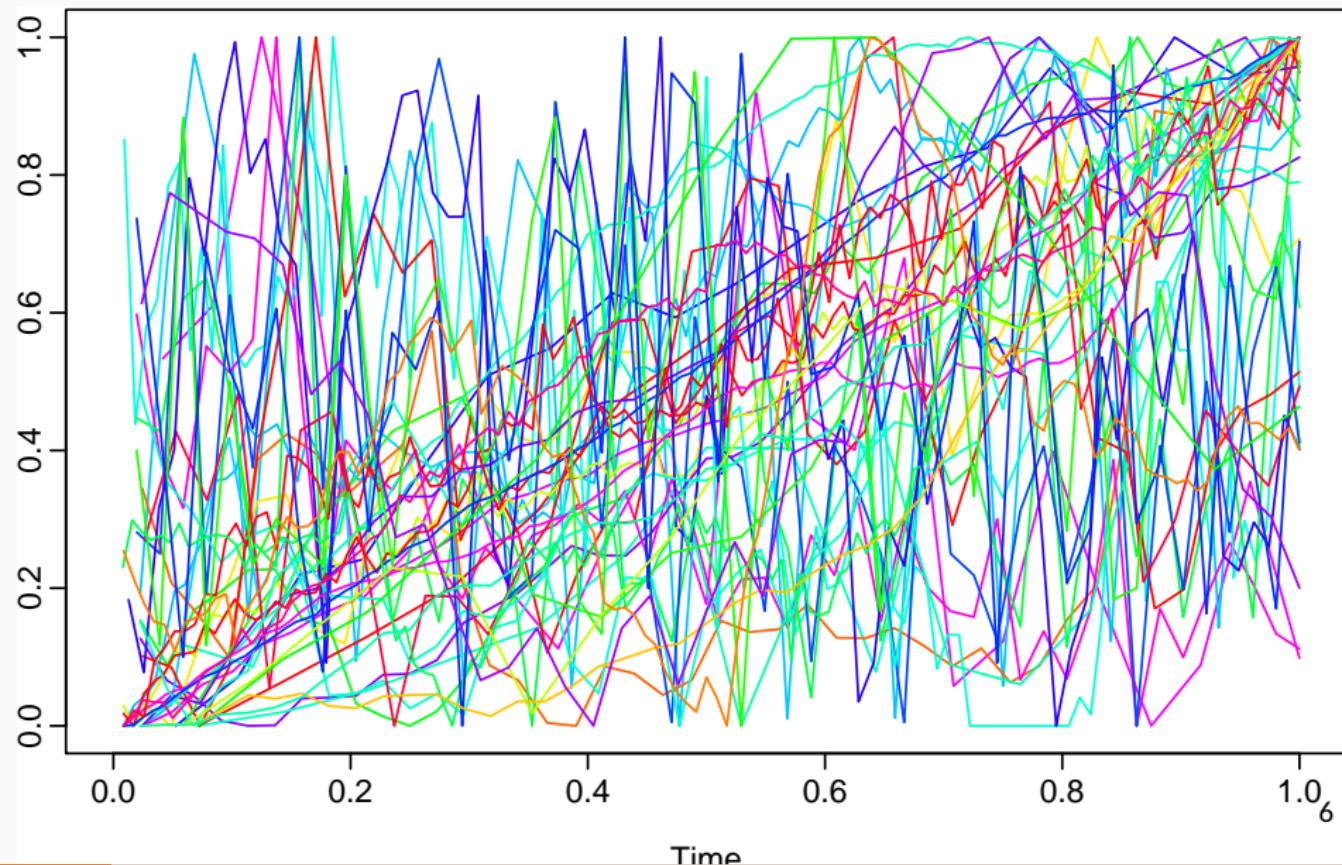
How to plot lots of time series?



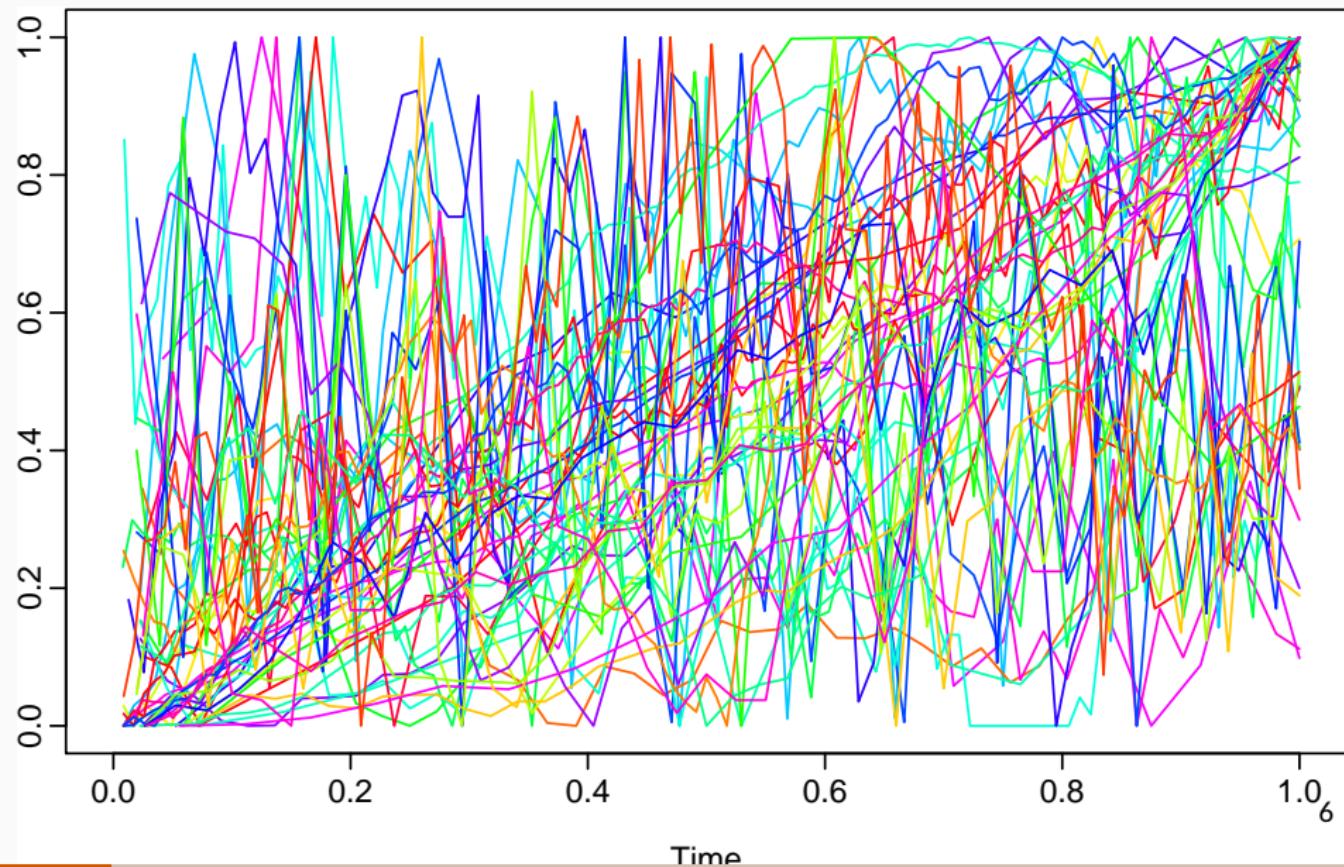
How to plot lots of time series?



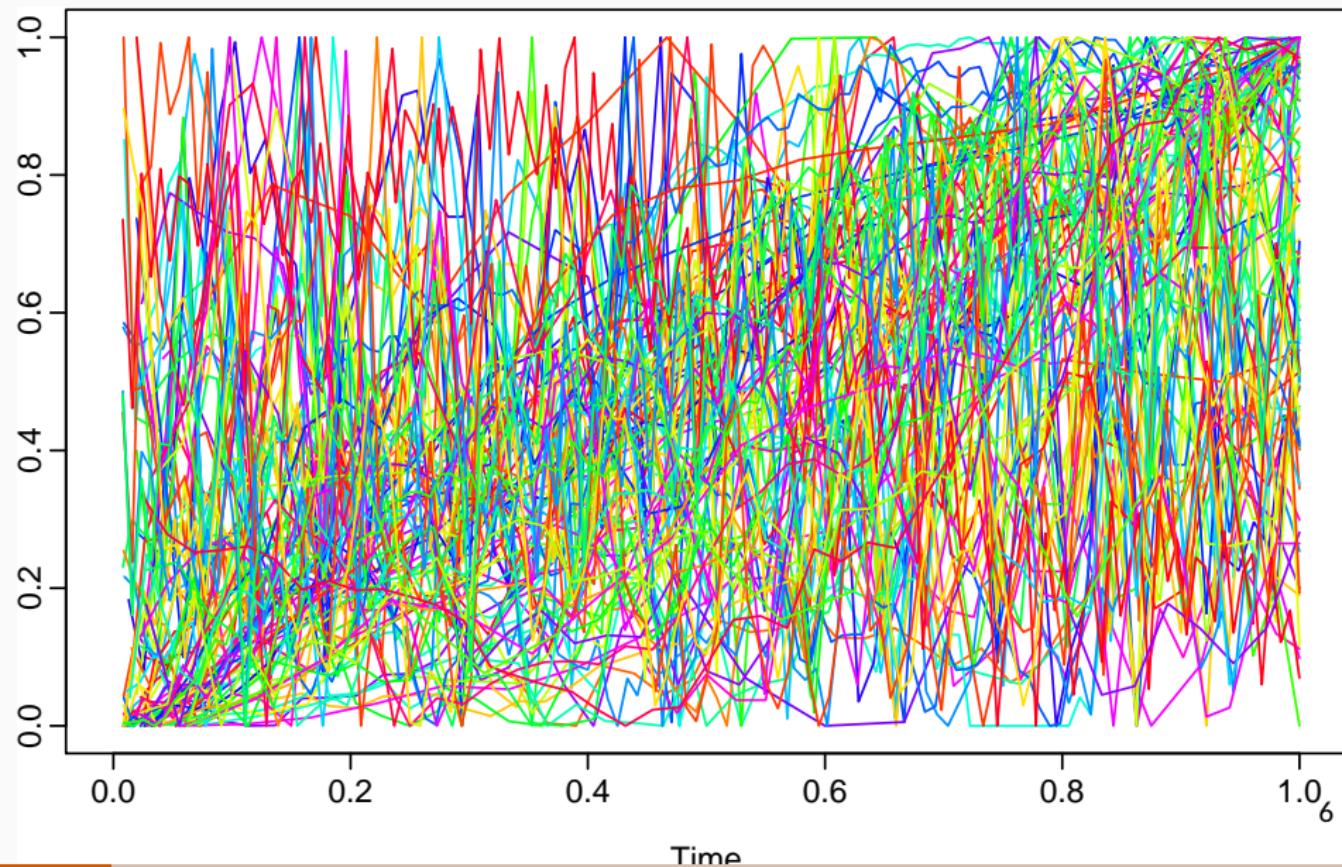
How to plot lots of time series?



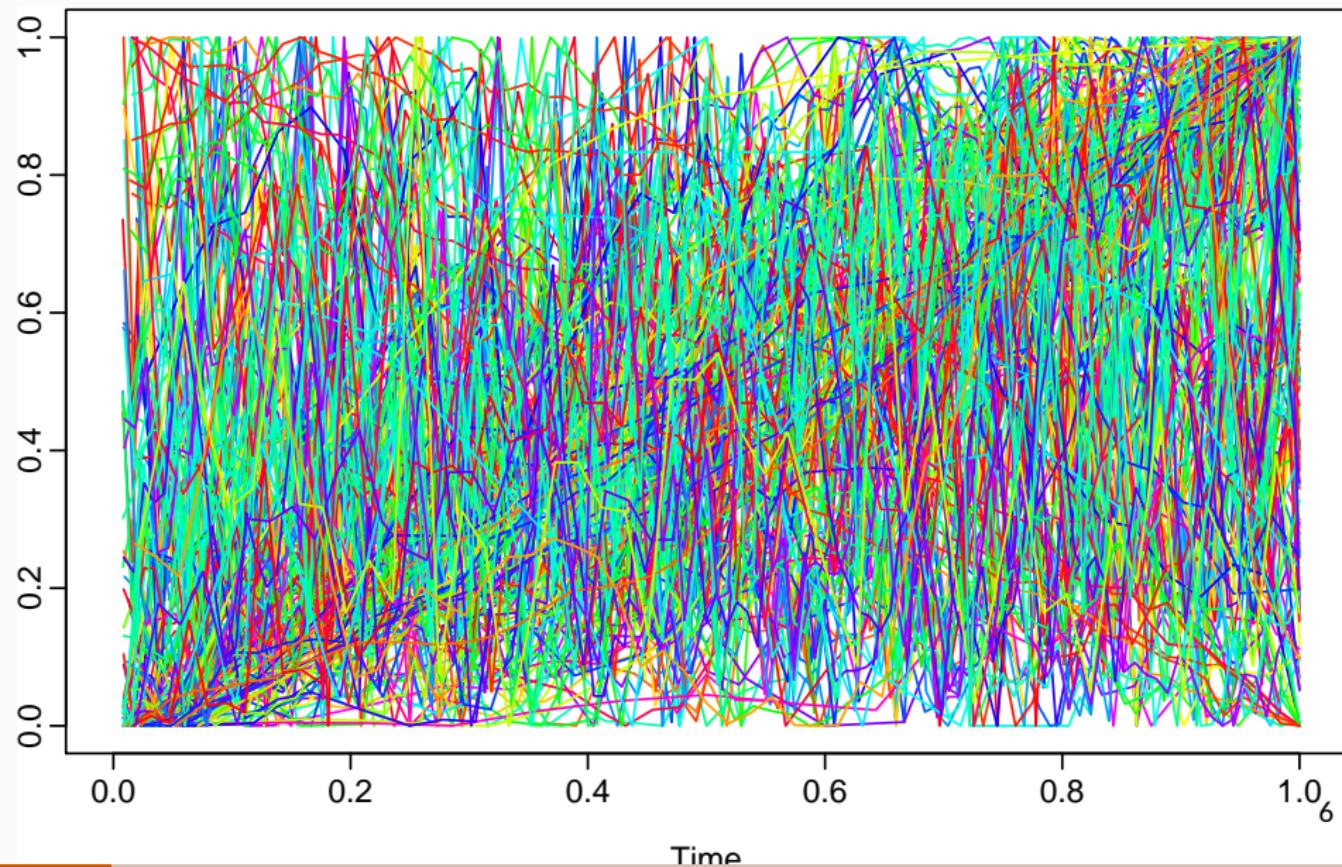
How to plot lots of time series?



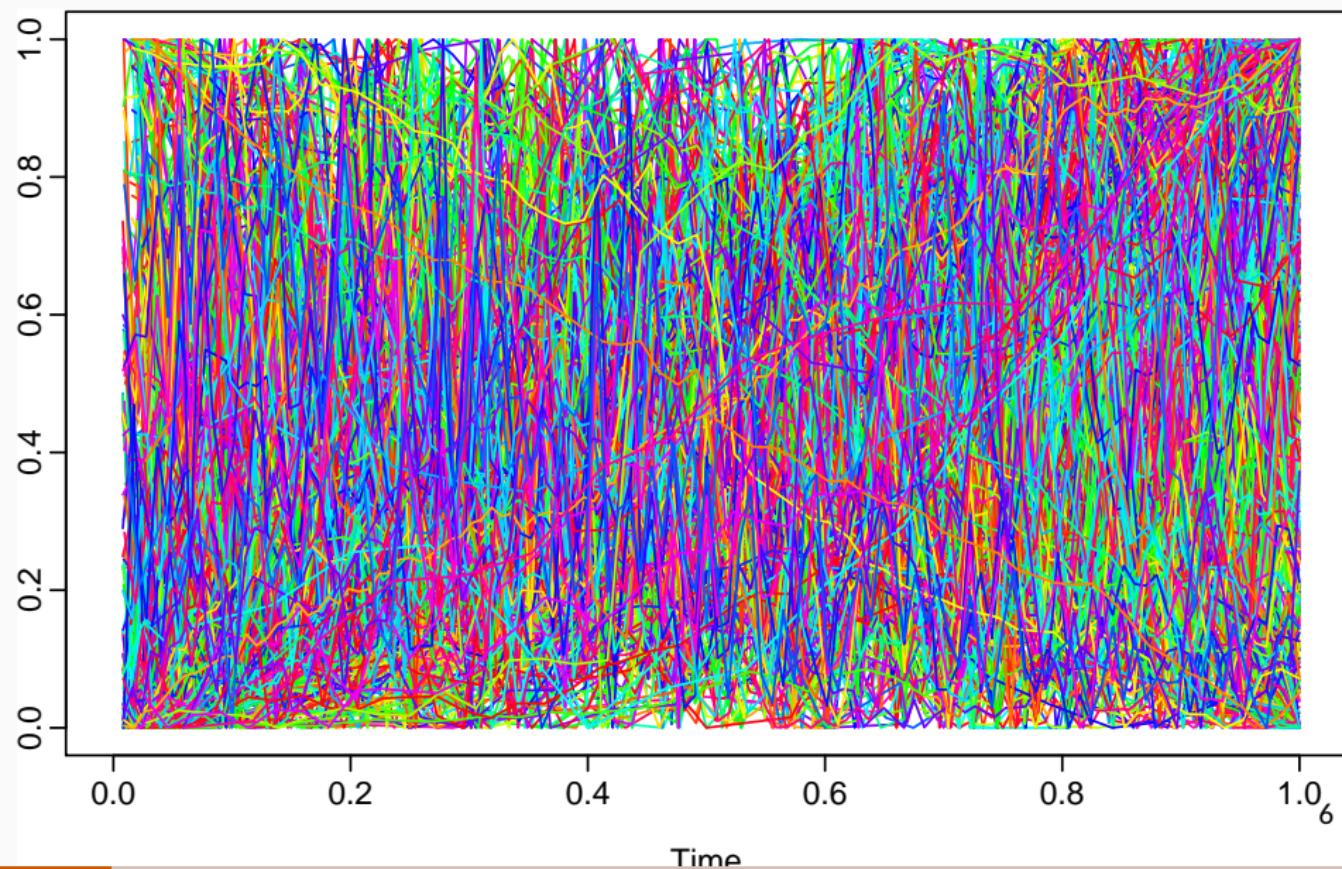
How to plot lots of time series?



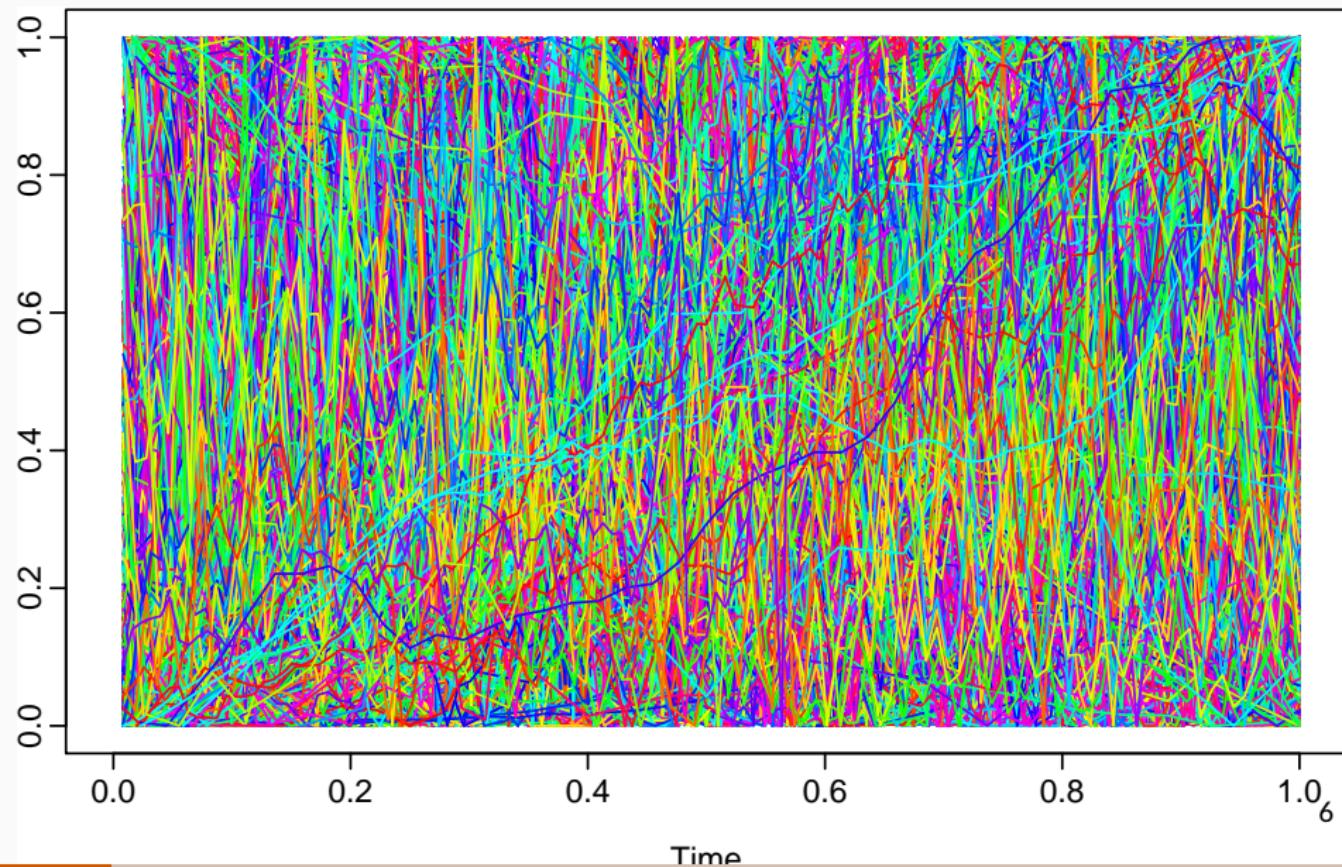
How to plot lots of time series?



How to plot lots of time series?



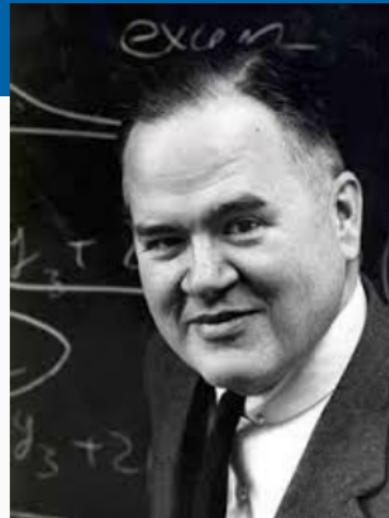
How to plot lots of time series?



Key idea

Cognostics

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

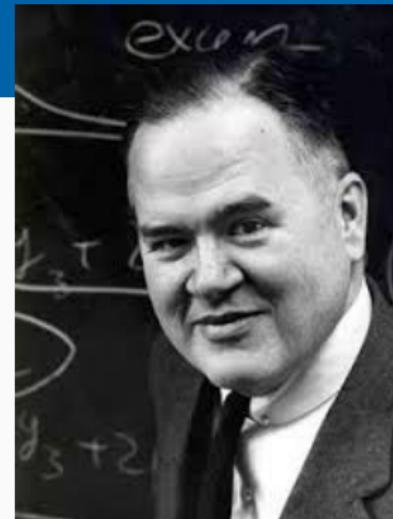


John W Tukey

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” in the machine learning literature.

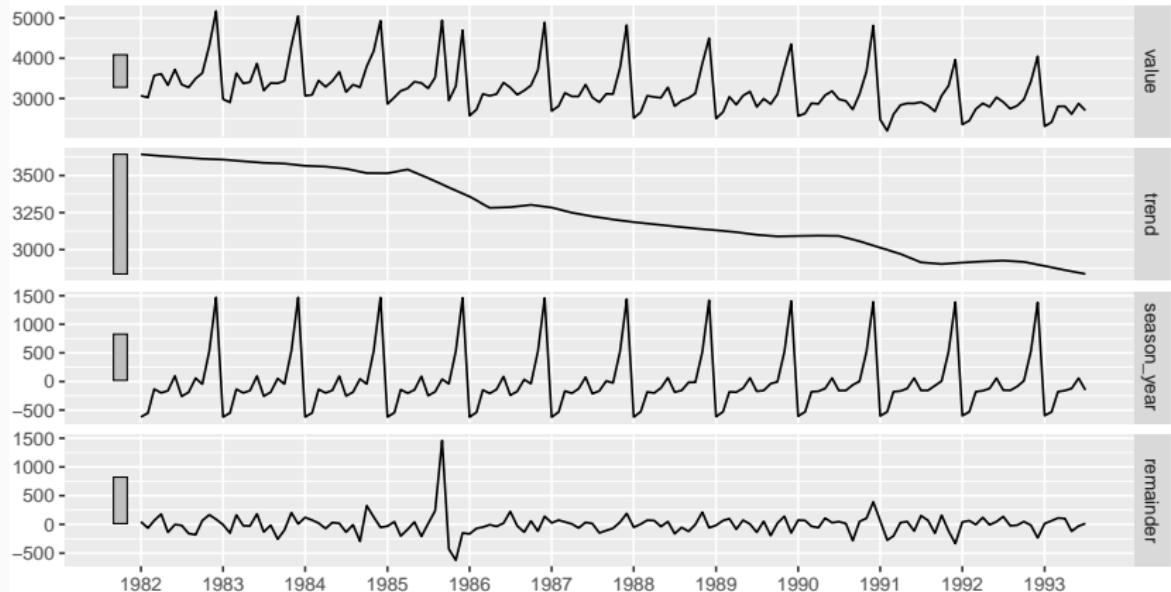
An STL decomposition: N2096

$$Y_t = S_t + T_t + R_t$$

S_t is periodic with mean 0

STL decomposition

value = trend + season_year + remainder



Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Autocorrelations of data (Y_1, \dots, Y_T)
- Autocorrelations of data (R_1, \dots, R_T)
- Strength of seasonality: $\max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)} \right)$
- Strength of trend: $\max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)} \right)$
- 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).
- Optimal Box-Cox transformation of data

Feature properties

For series with different lengths, scales, domains, etc., we need features that are:

- scale-independent
- ergodic

Feature properties

For series with different lengths, scales, domains, etc., we need features that are:

- scale-independent
- ergodic

- 1 Seasonal period
- 2 Strength of seasonality
- 3 Strength of trend
- 4 First autocorrelation of STL remainder series
- 5 Spectral entropy
- 6 Optimal MLE Box-Cox transformation of data

Feature properties

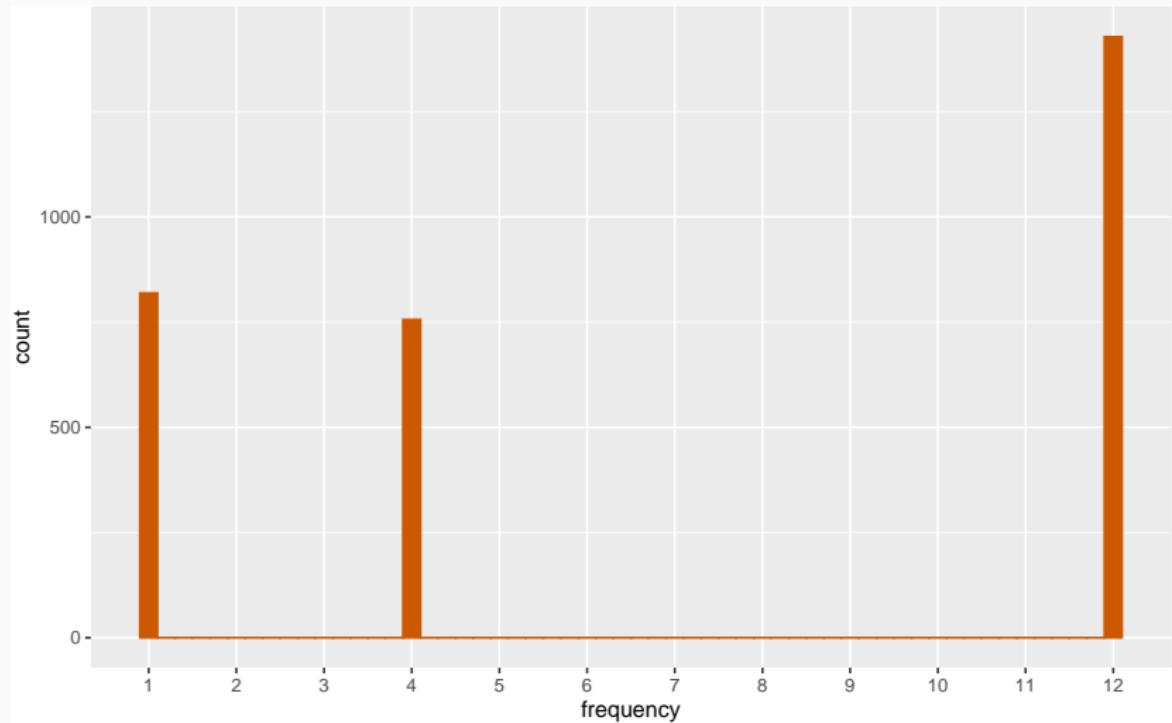
For series with different lengths, scales, domains, etc., we need features that are:

- scale-independent
- ergodic

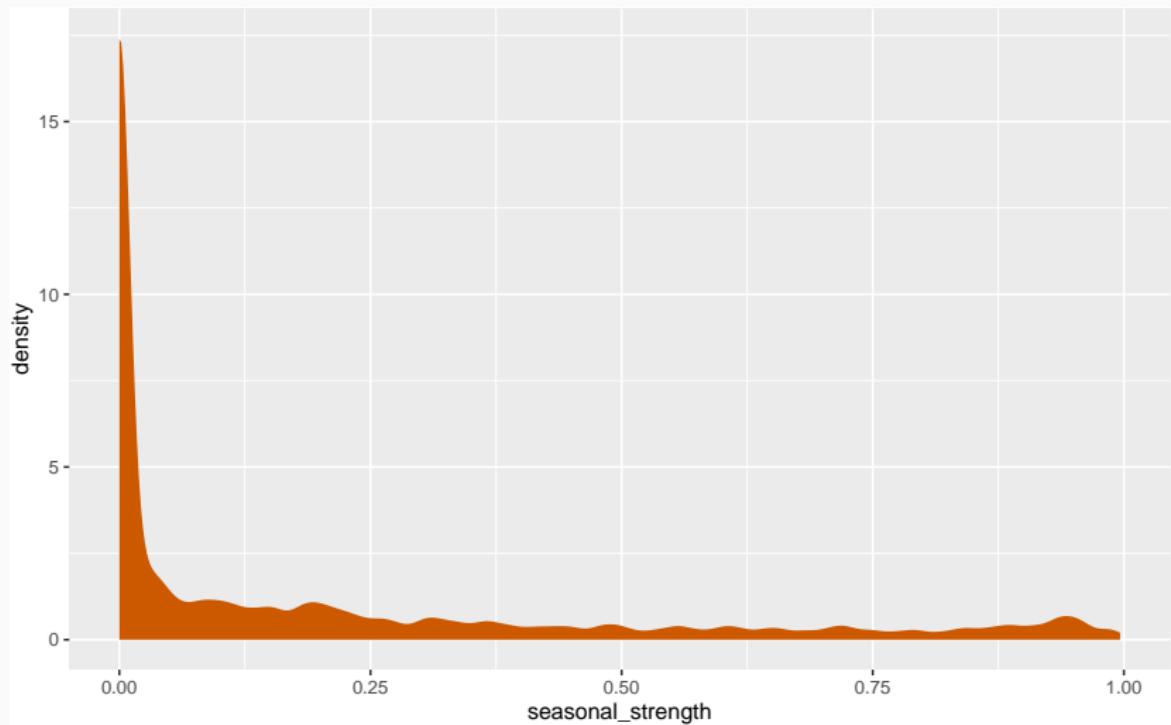
- 1 Seasonal period
- 2 Strength of seasonality
- 3 Strength of trend
- 4 First autocorrelation of STL remainder series
- 5 Spectral entropy
- 6 Optimal MLE Box-Cox transformation of data

For other analyses, it may be appropriate to have different requirements.

Distribution of Period for M3

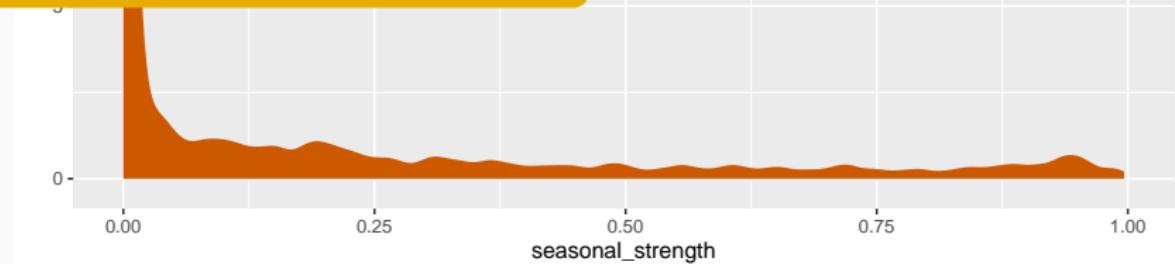
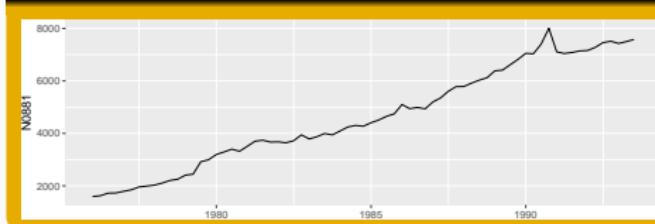


Distribution of Seasonality for M3

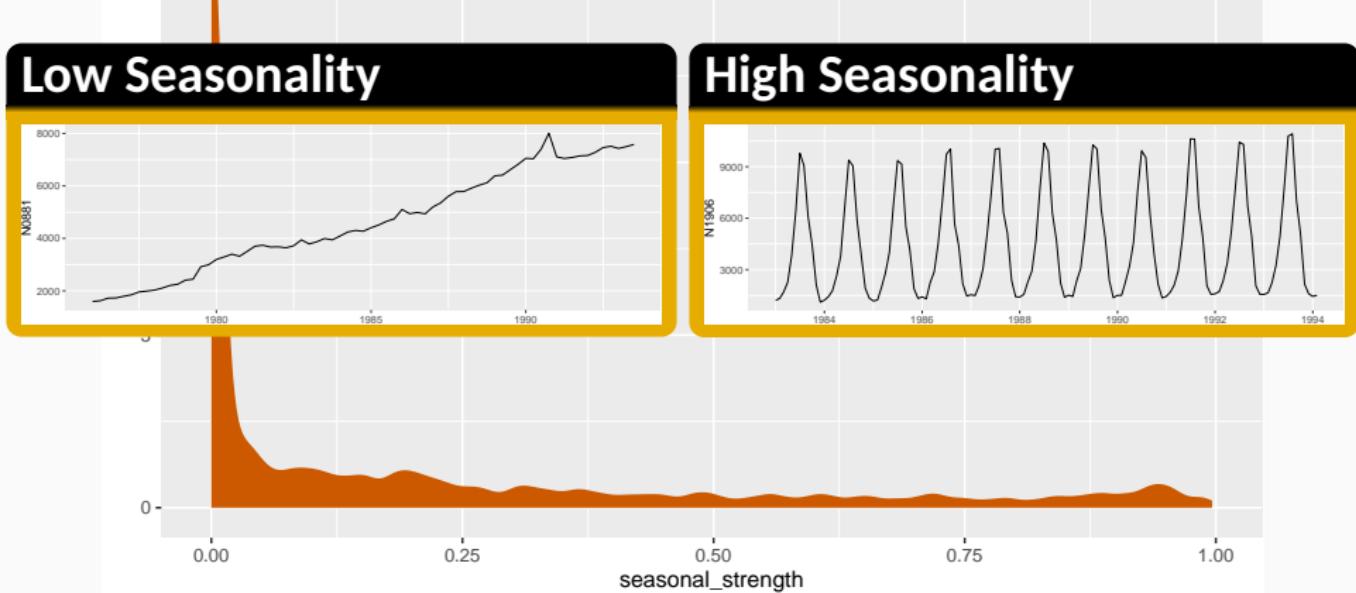


Distribution of Seasonality for M3

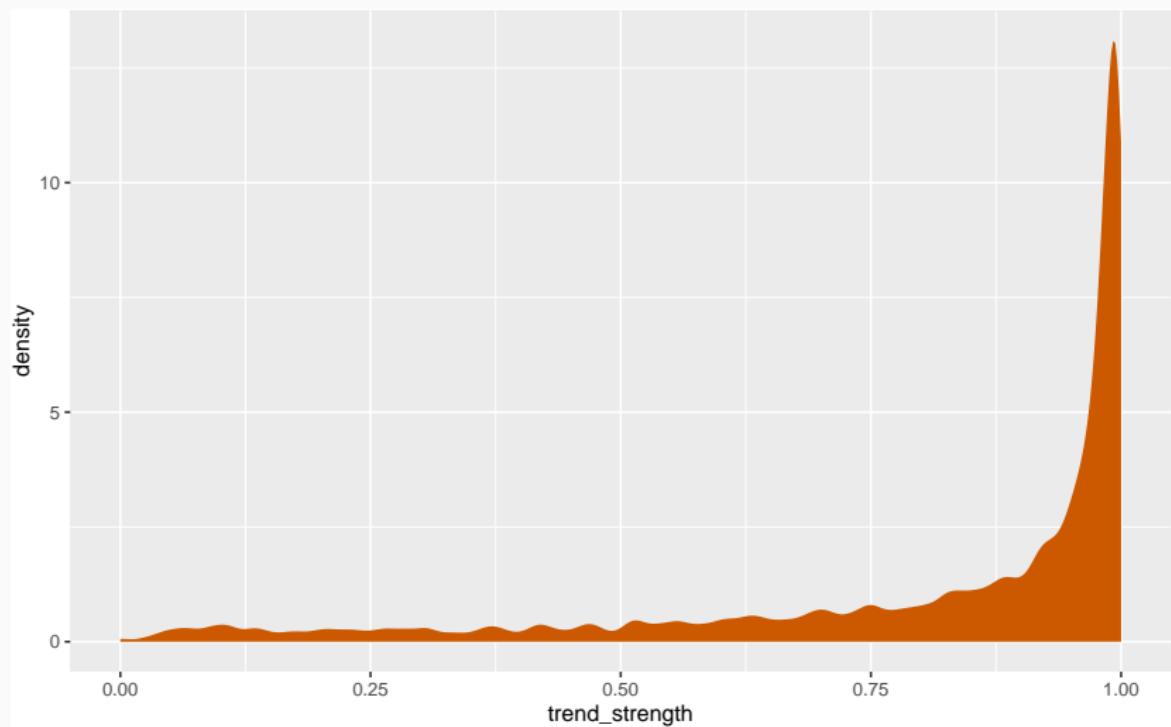
Low Seasonality



Distribution of Seasonality for M3

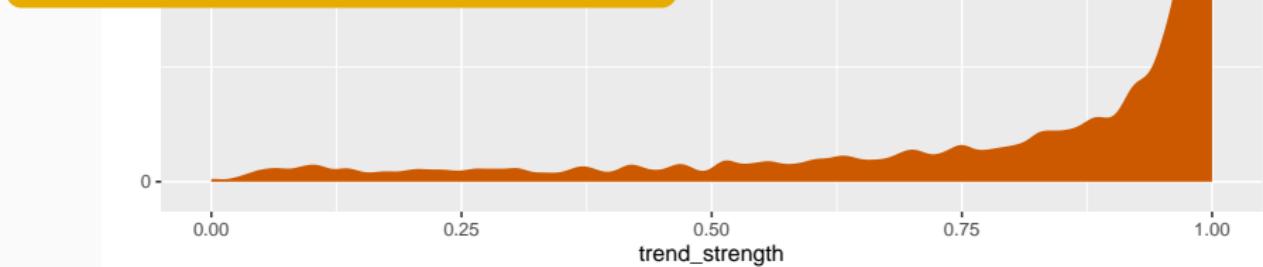
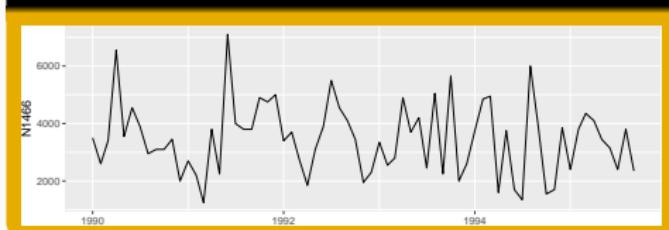


Distribution of Trend for M3

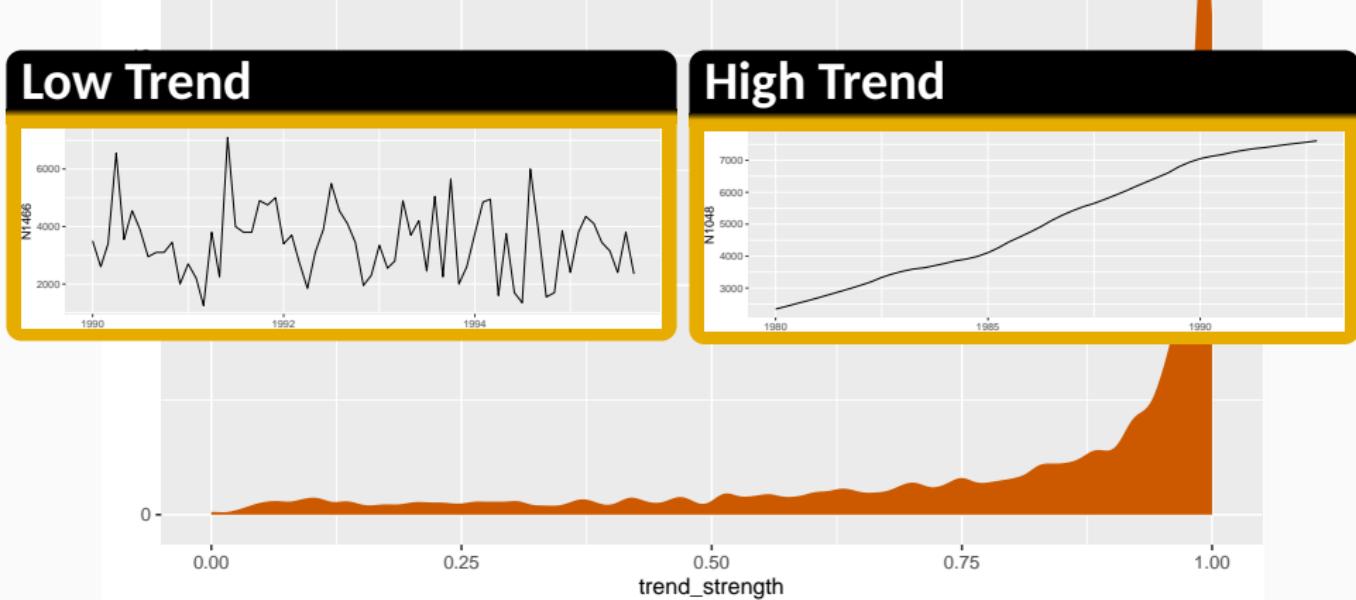


Distribution of Trend for M3

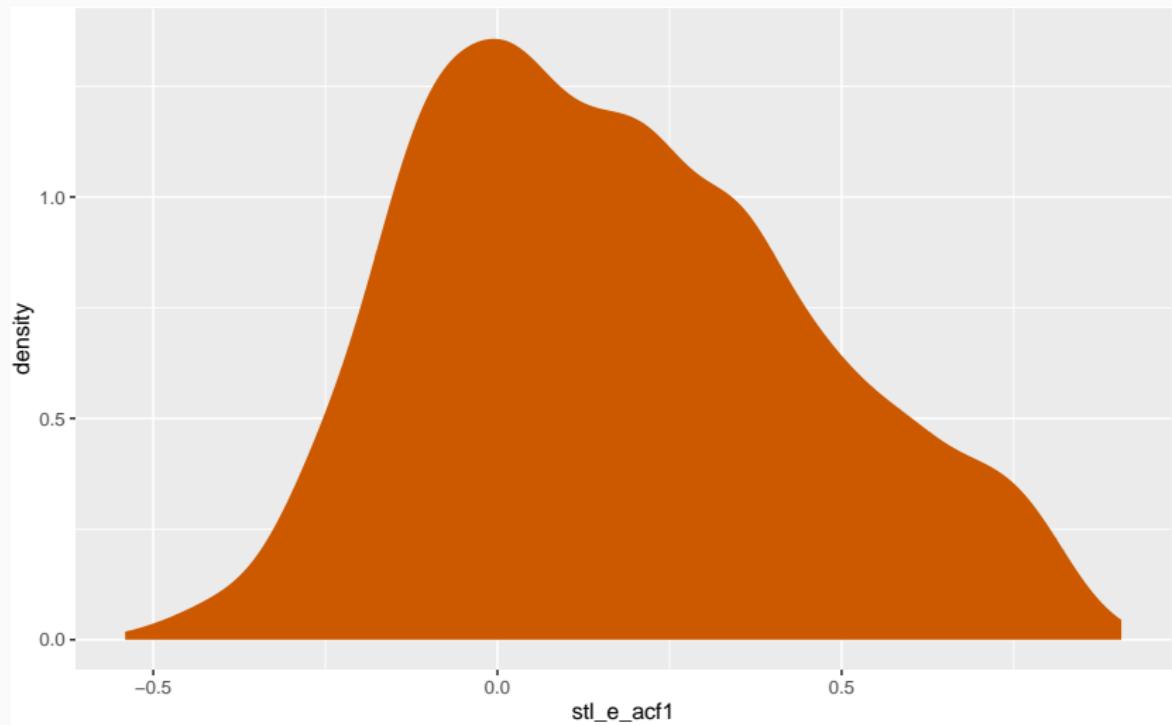
Low Trend



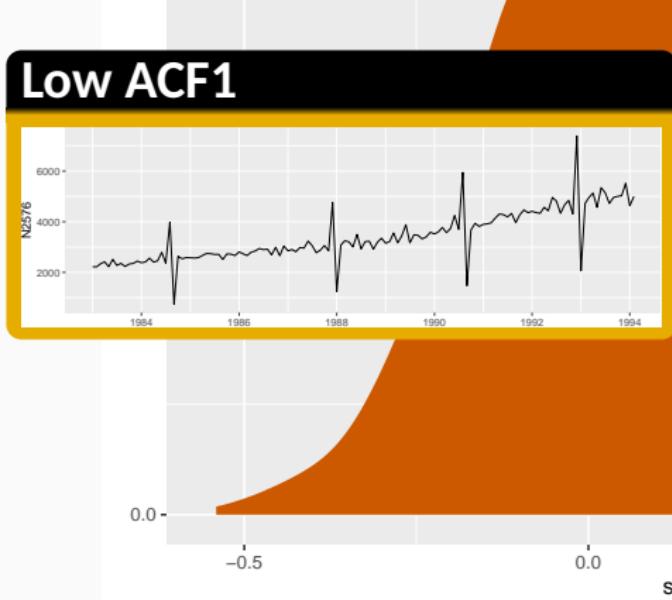
Distribution of Trend for M3



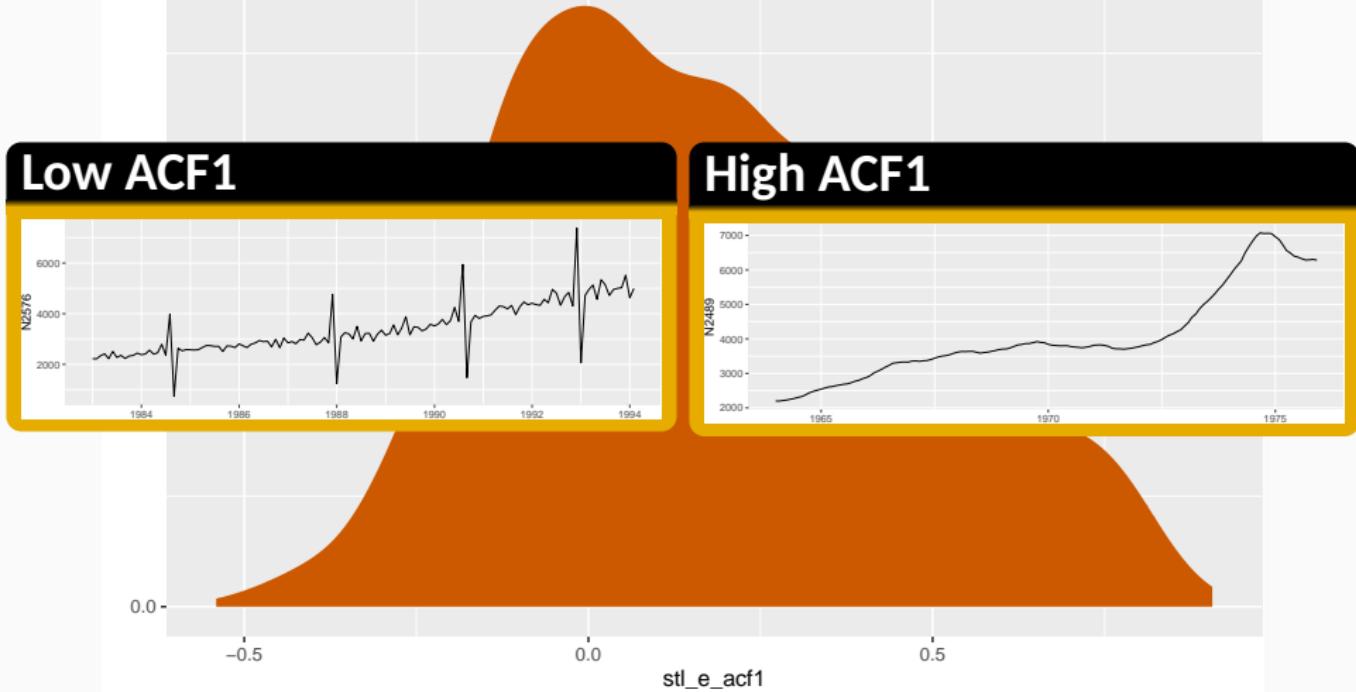
Distribution of Residual ACF1 for M3



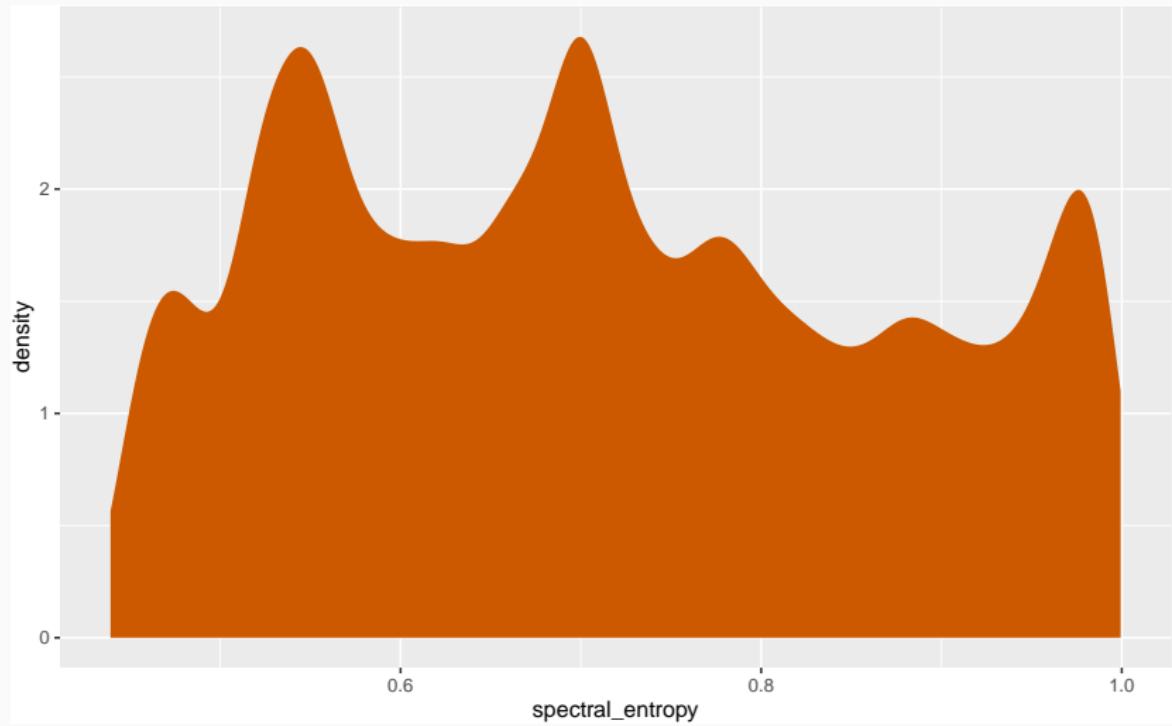
Distribution of Residual ACF1 for M3



Distribution of Residual ACF1 for M3

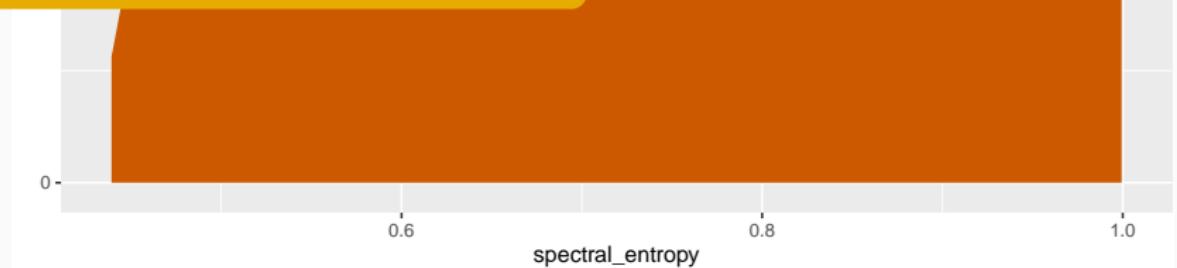
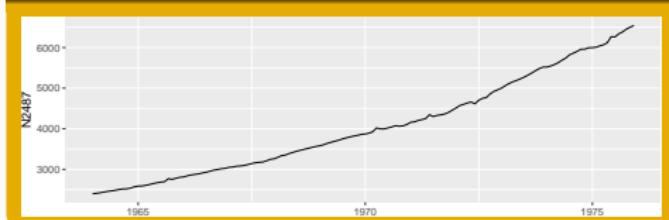


Distribution of Spectral Entropy for M3

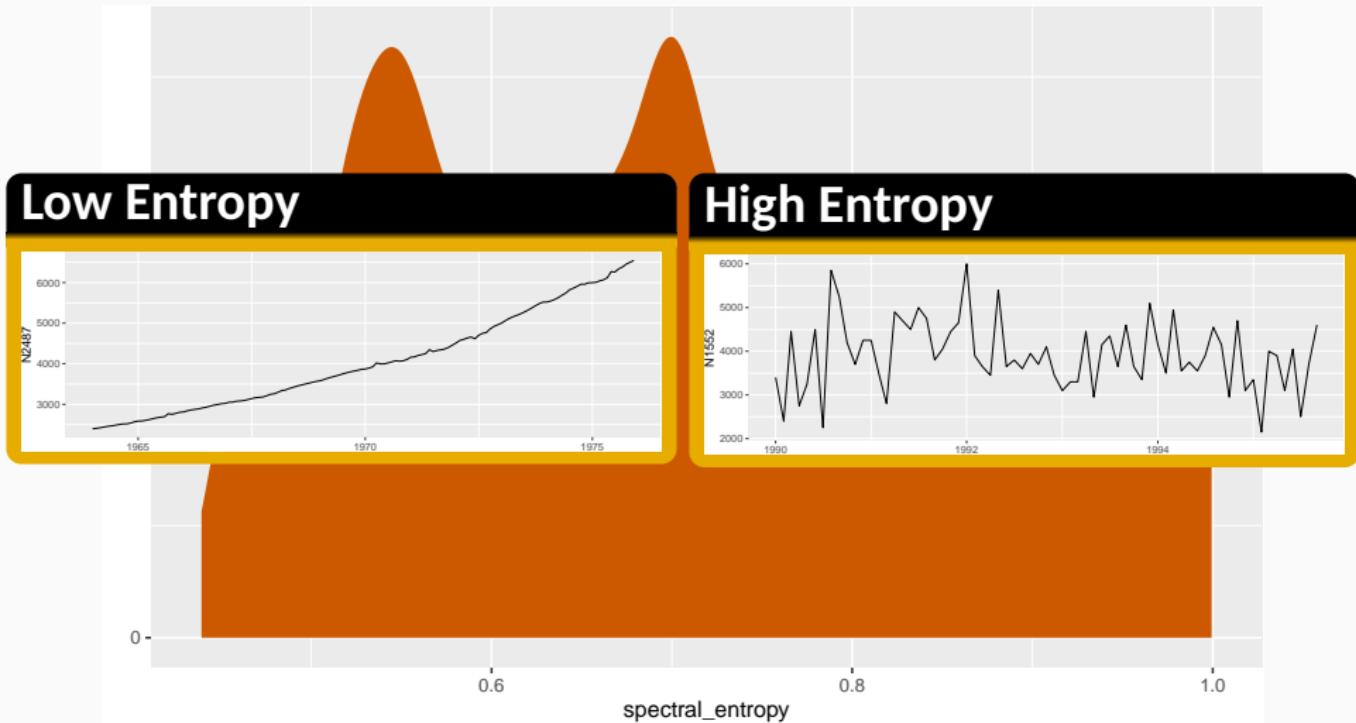


Distribution of Spectral Entropy for M3

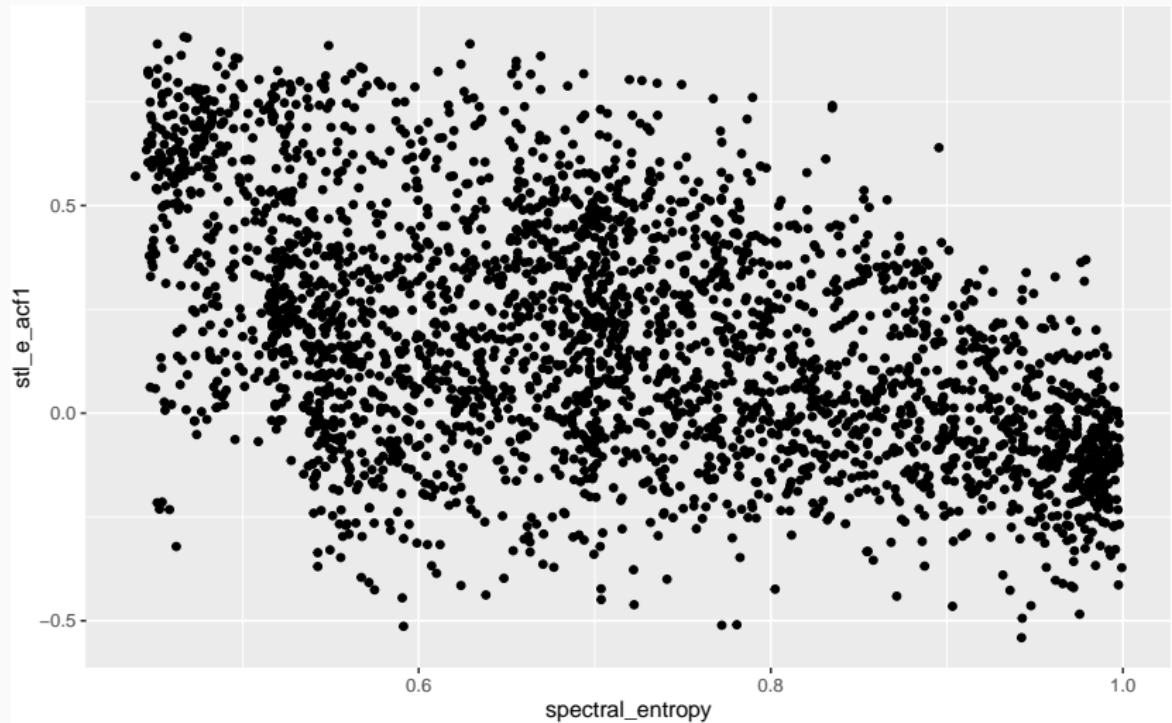
Low Entropy



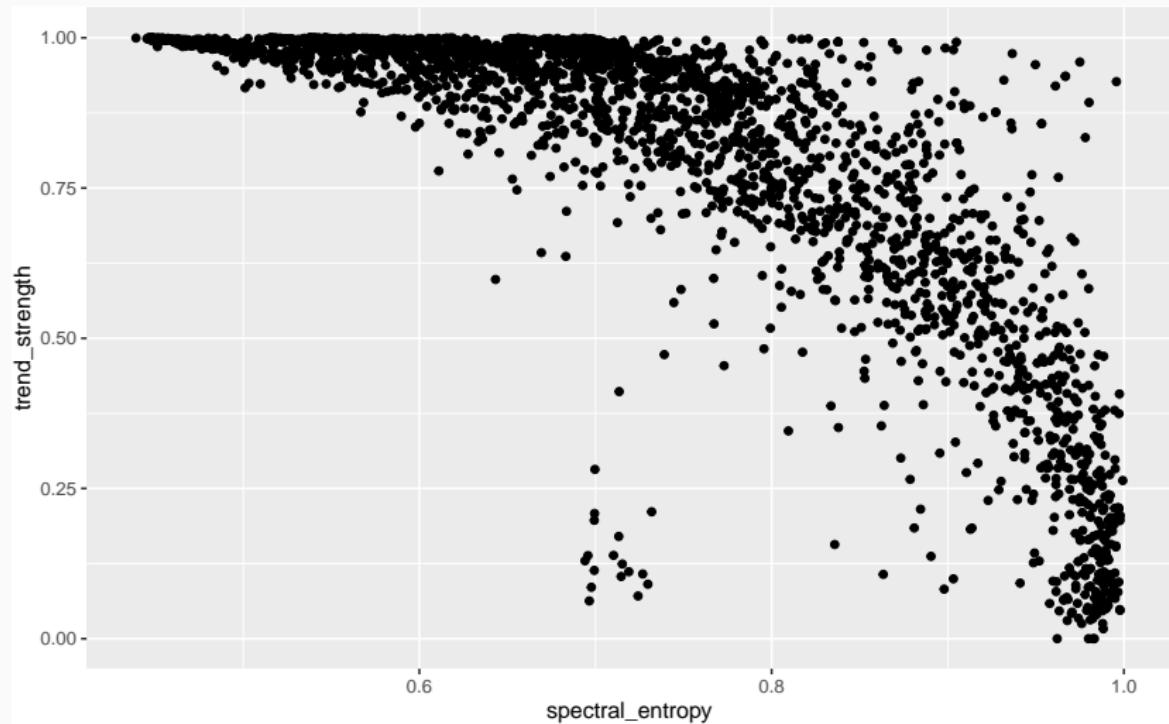
Distribution of Spectral Entropy for M3



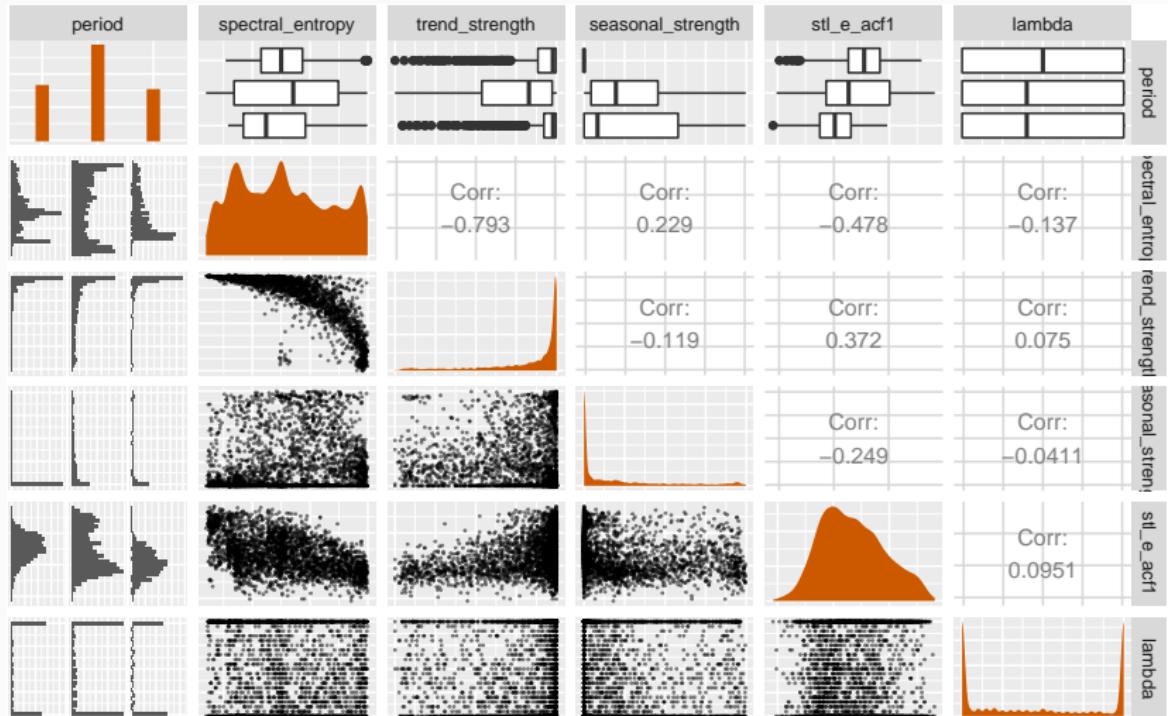
Feature distributions



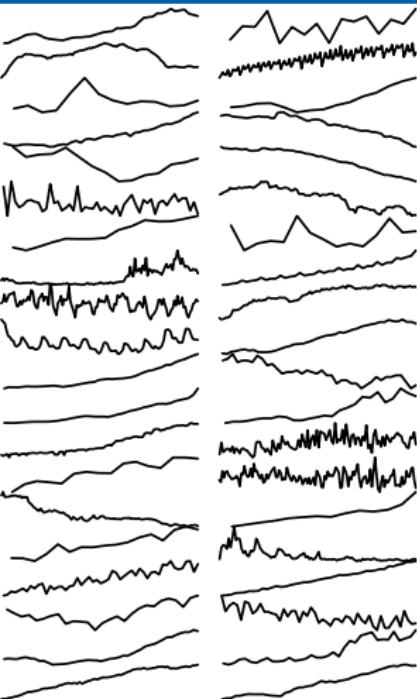
Feature distributions



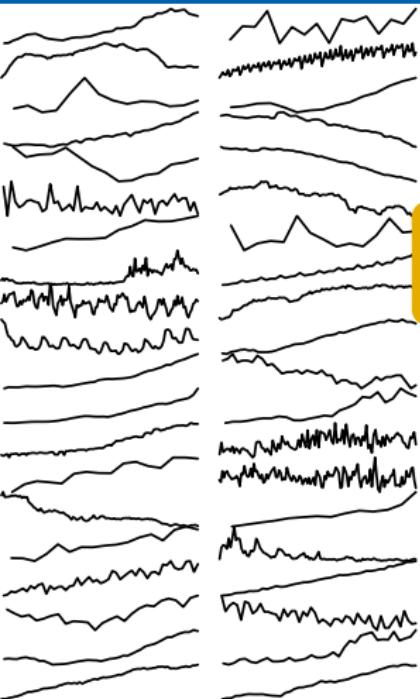
Feature distributions



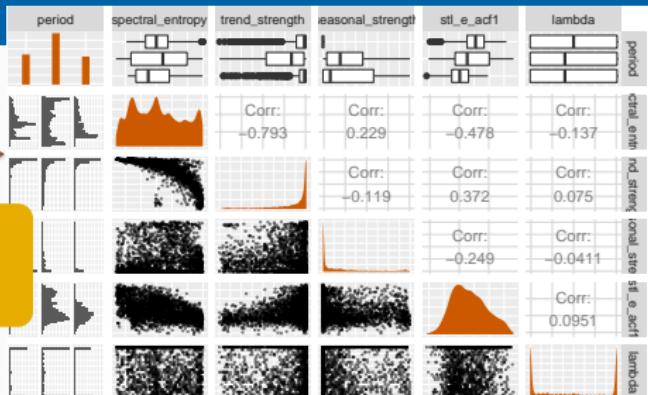
Dimension reduction for time series



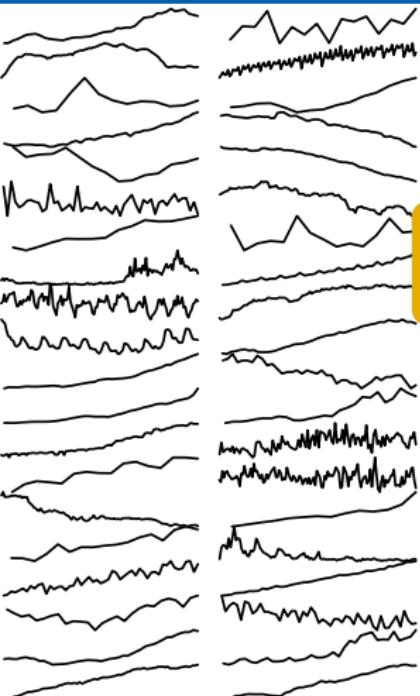
Dimension reduction for time series



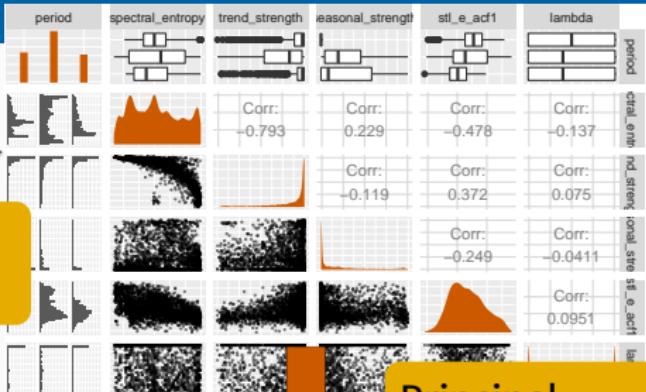
Feature
calculation



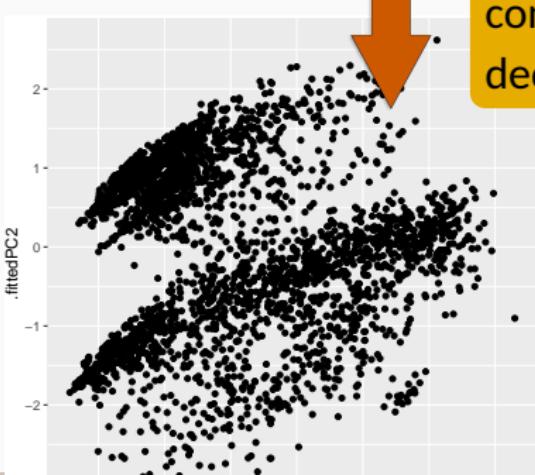
Dimension reduction for time series



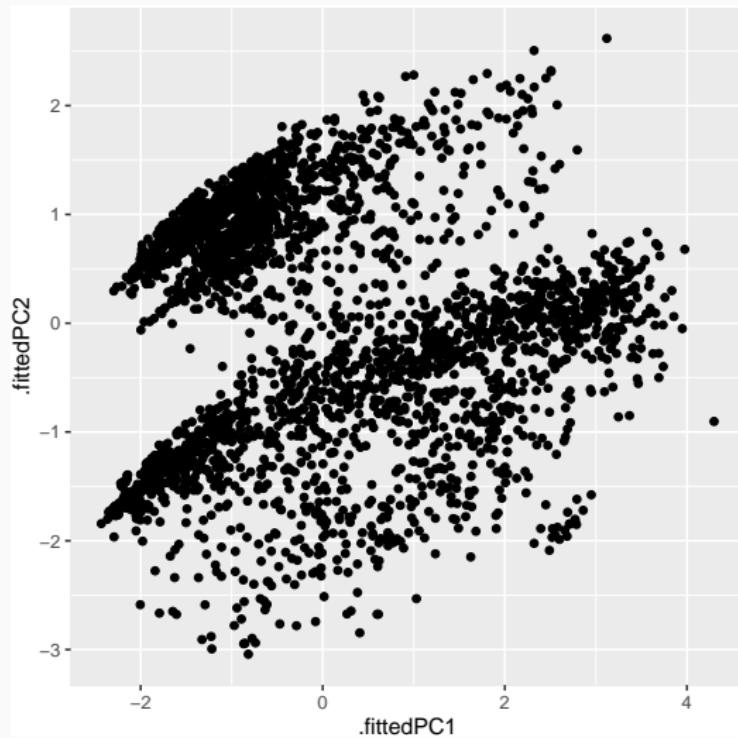
Feature
calculation



Principal
component
decomposition

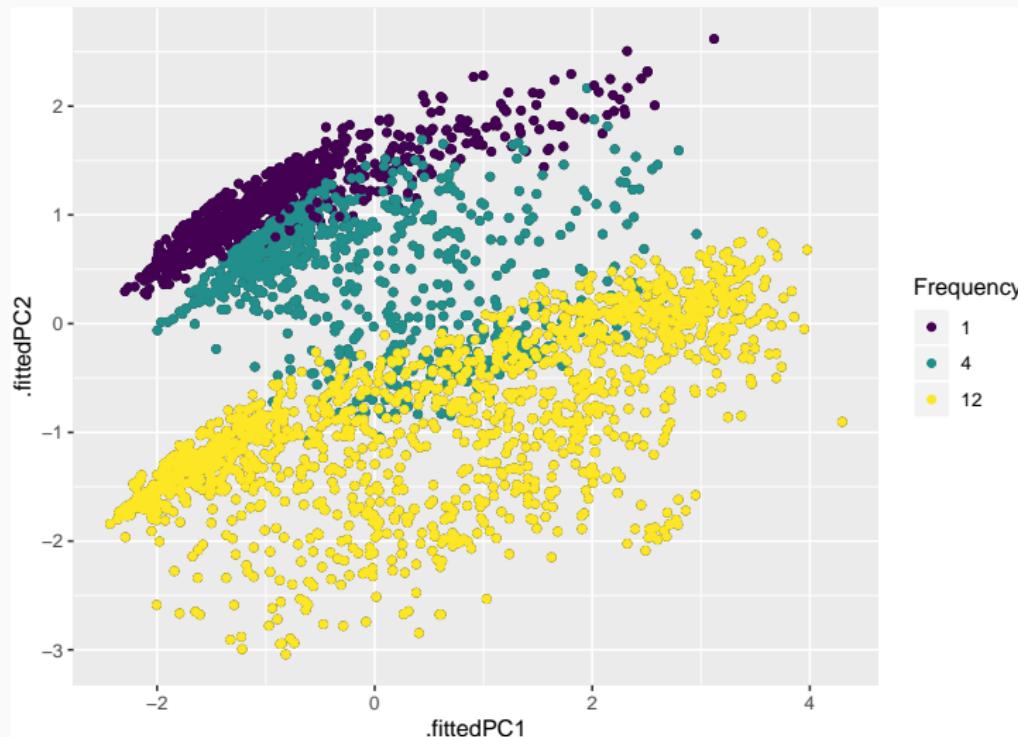


M3 feature space

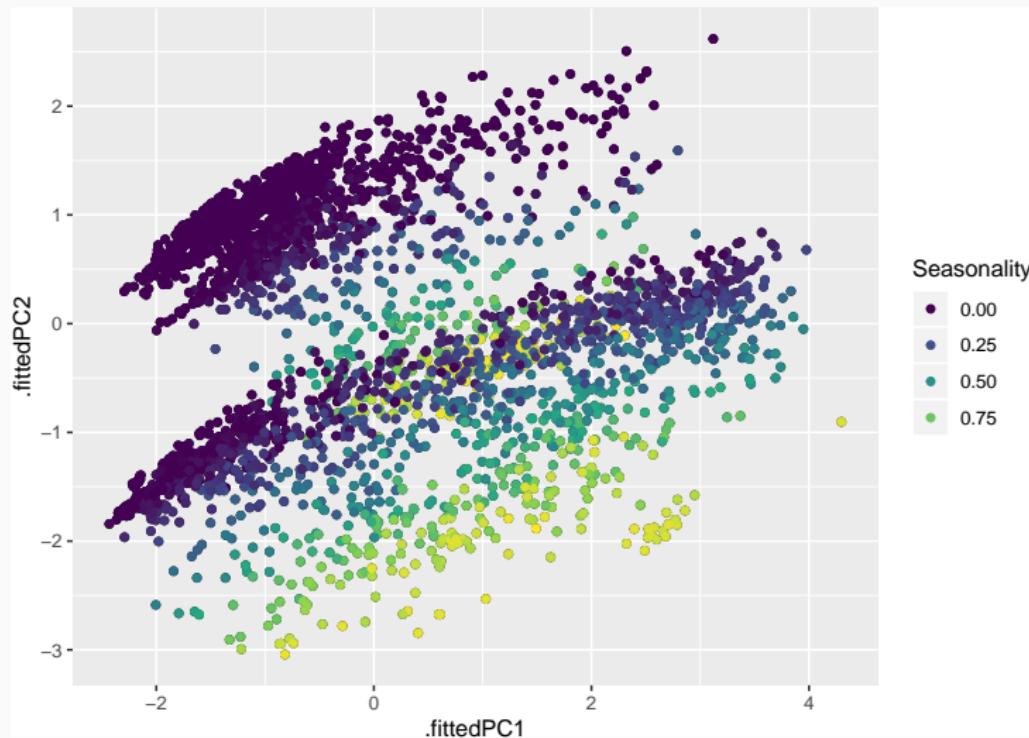


First two PCs explain
58.5% of the variance.

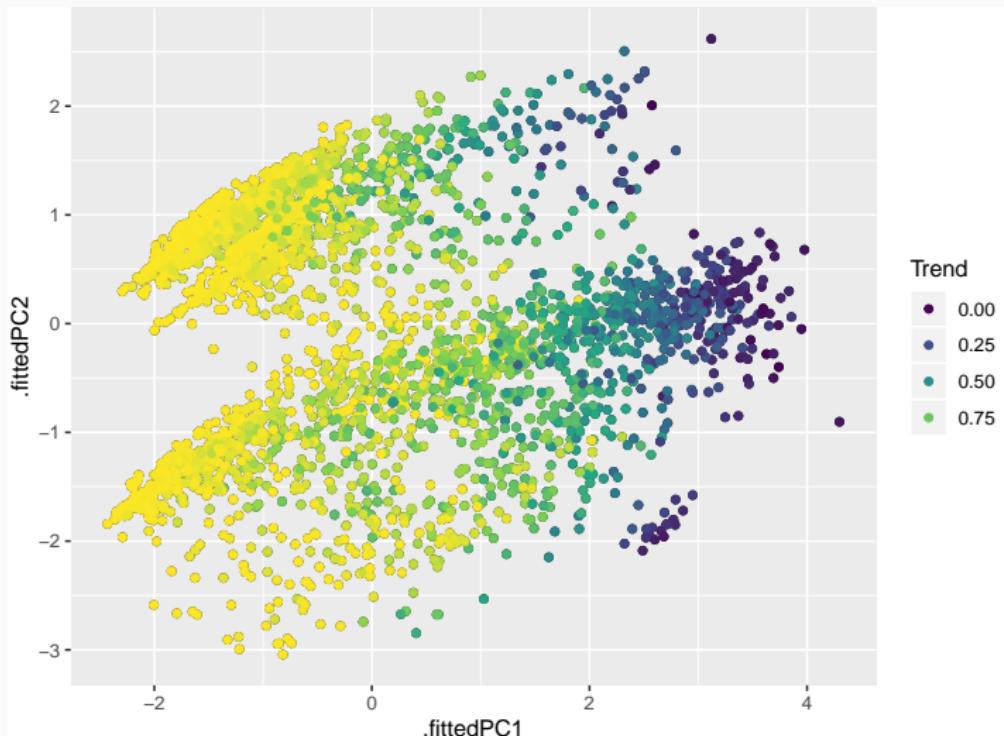
M3 feature space



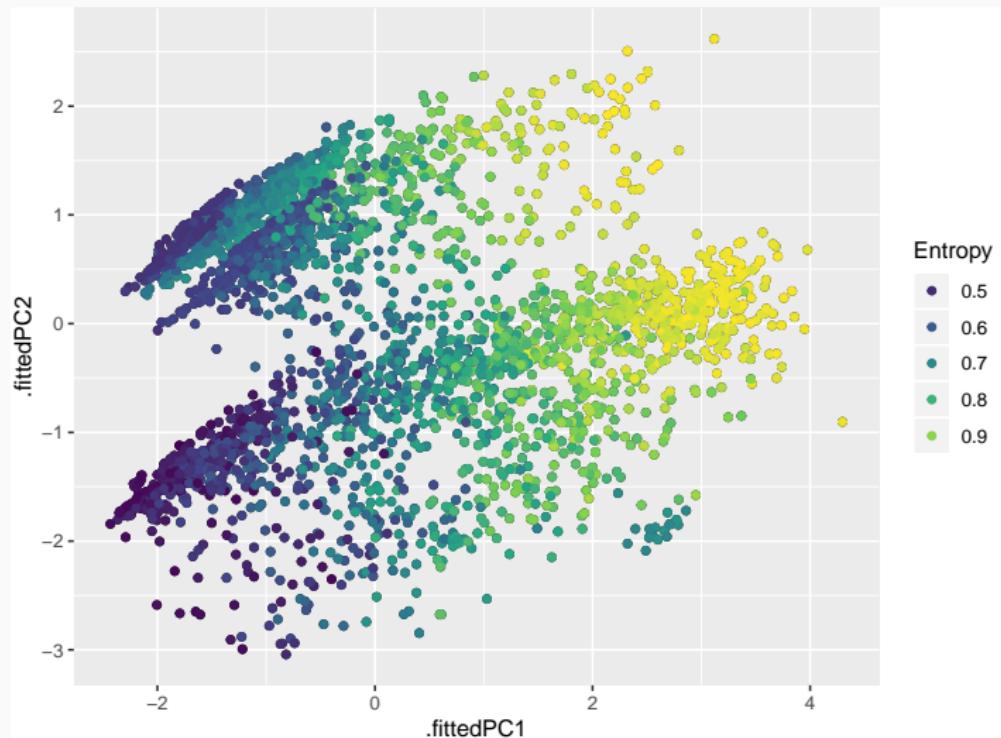
M3 feature space



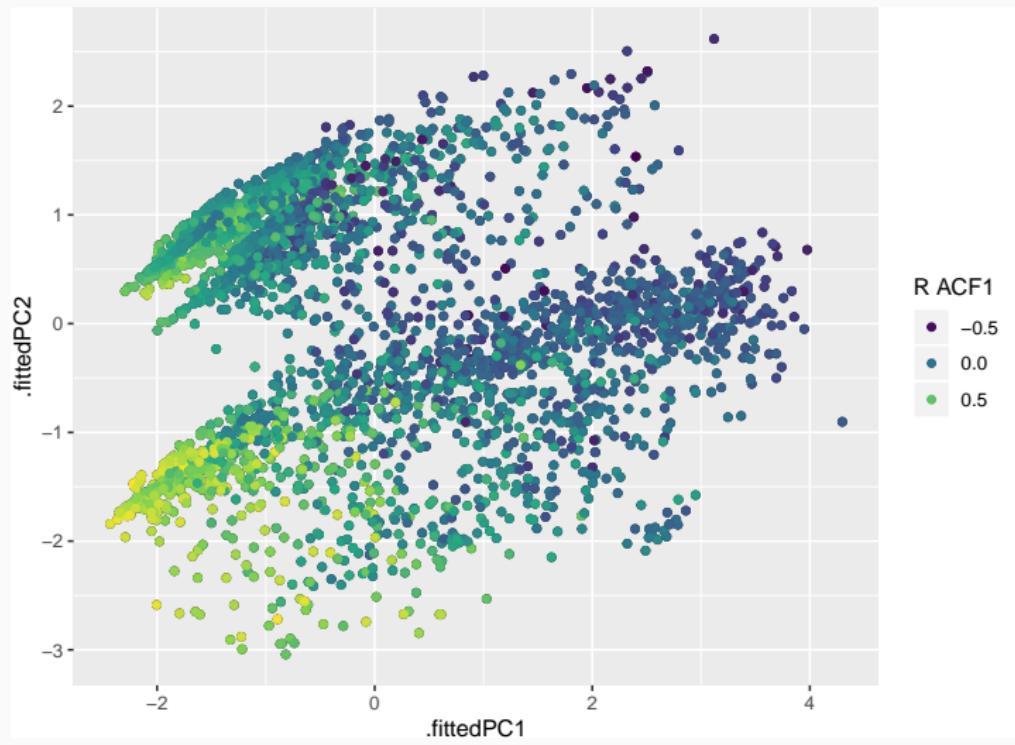
M3 feature space



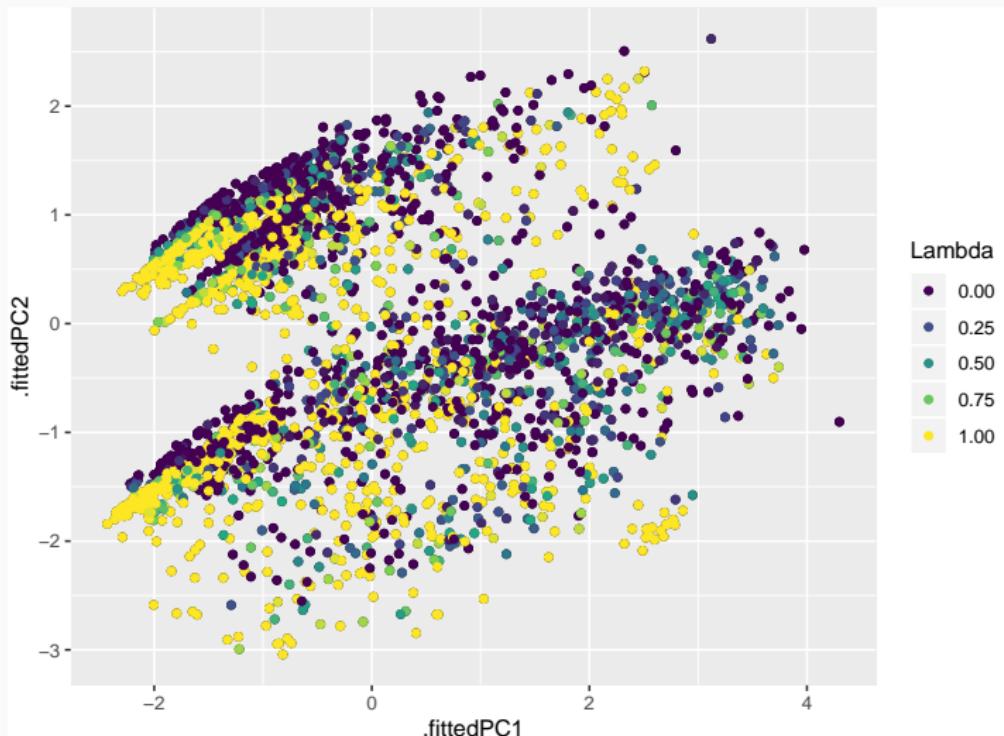
M3 feature space



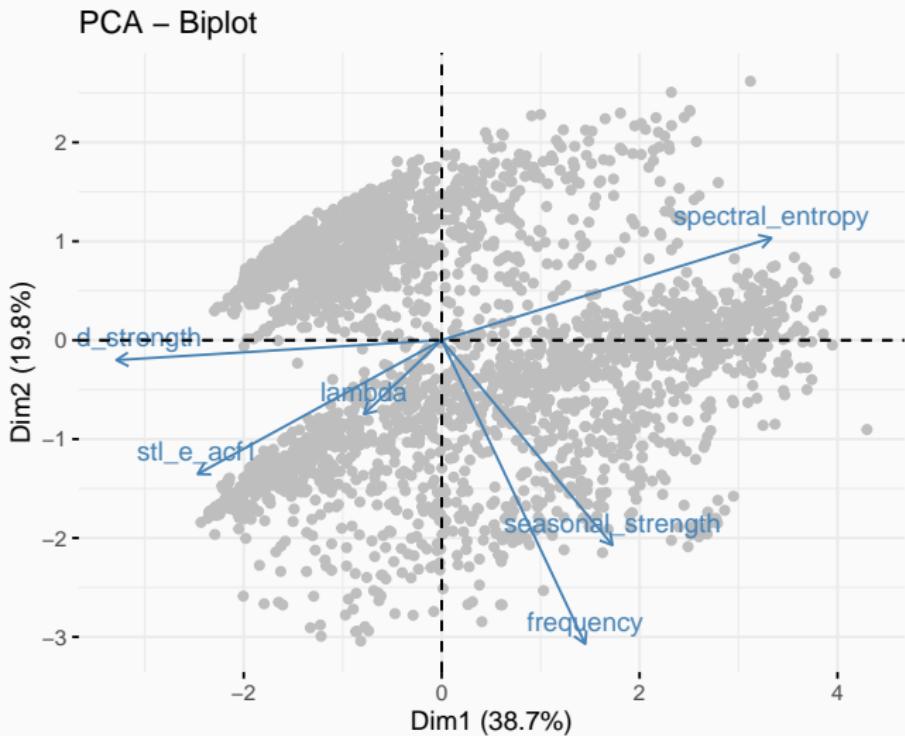
M3 feature space



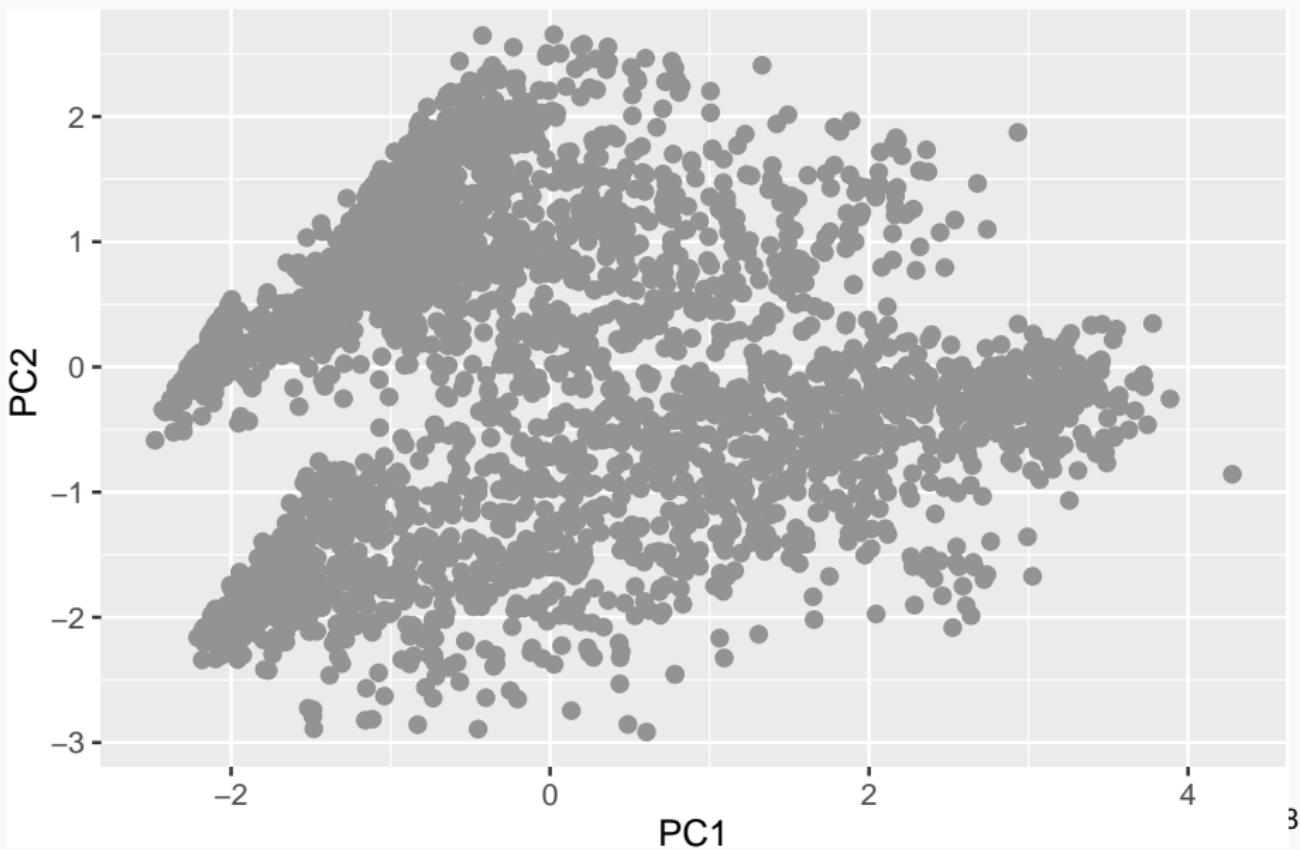
M3 feature space



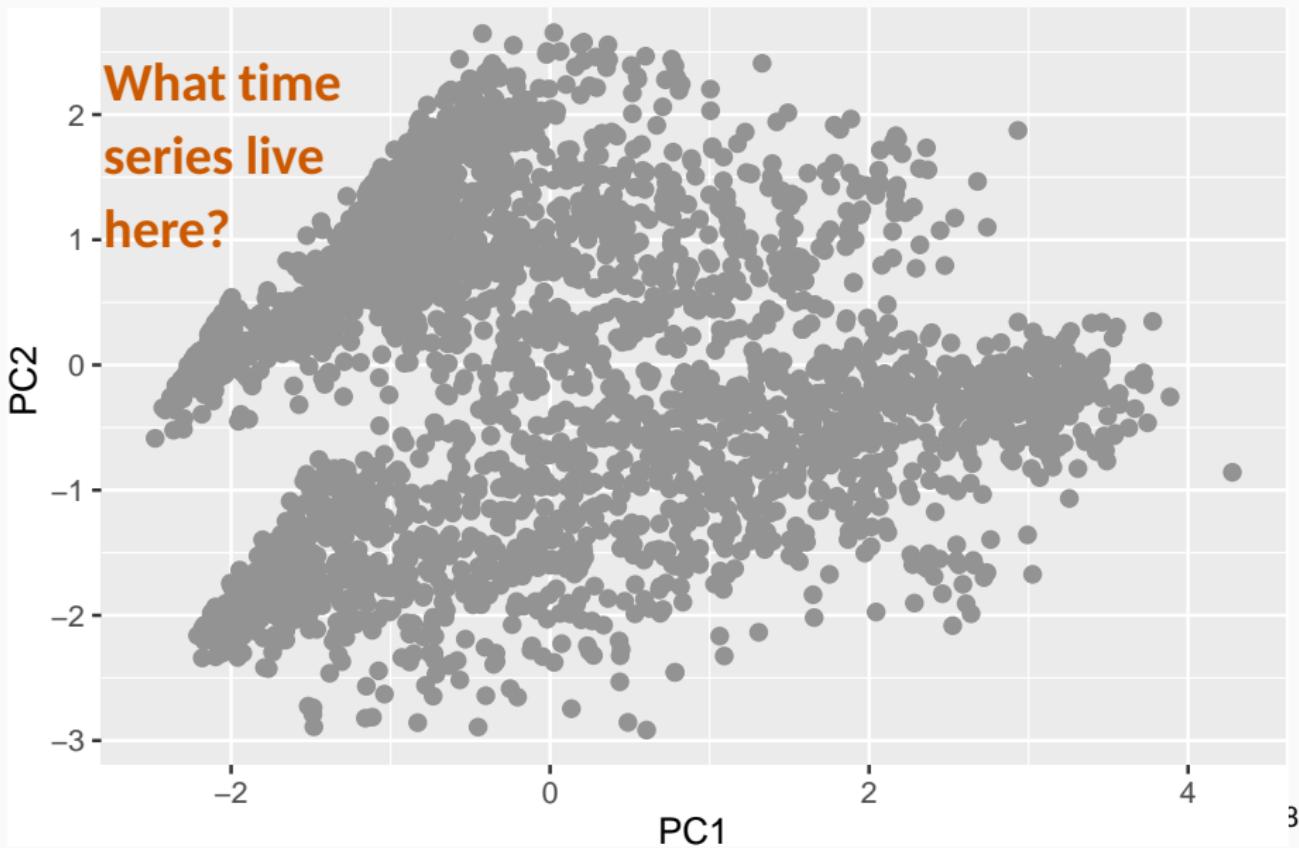
M3 feature space



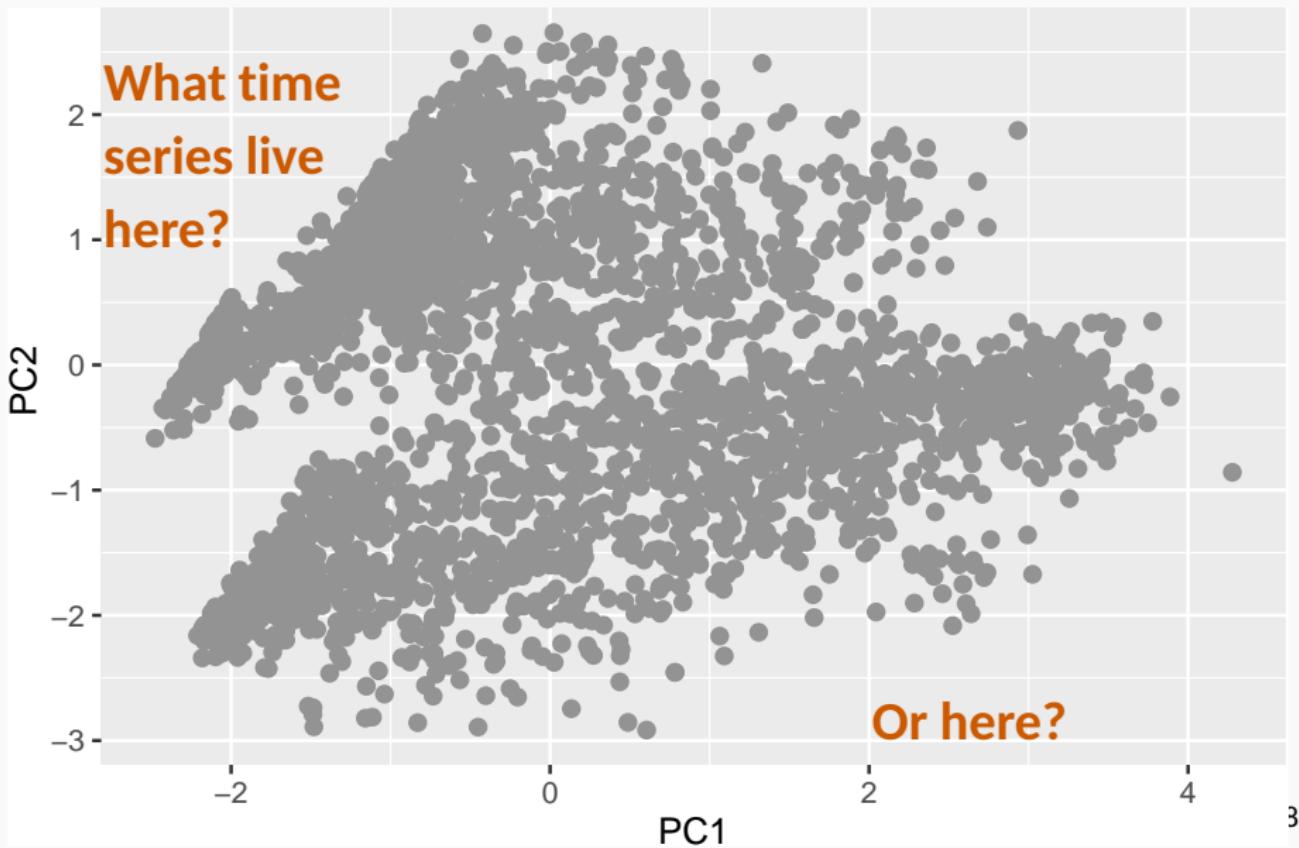
What about the holes?



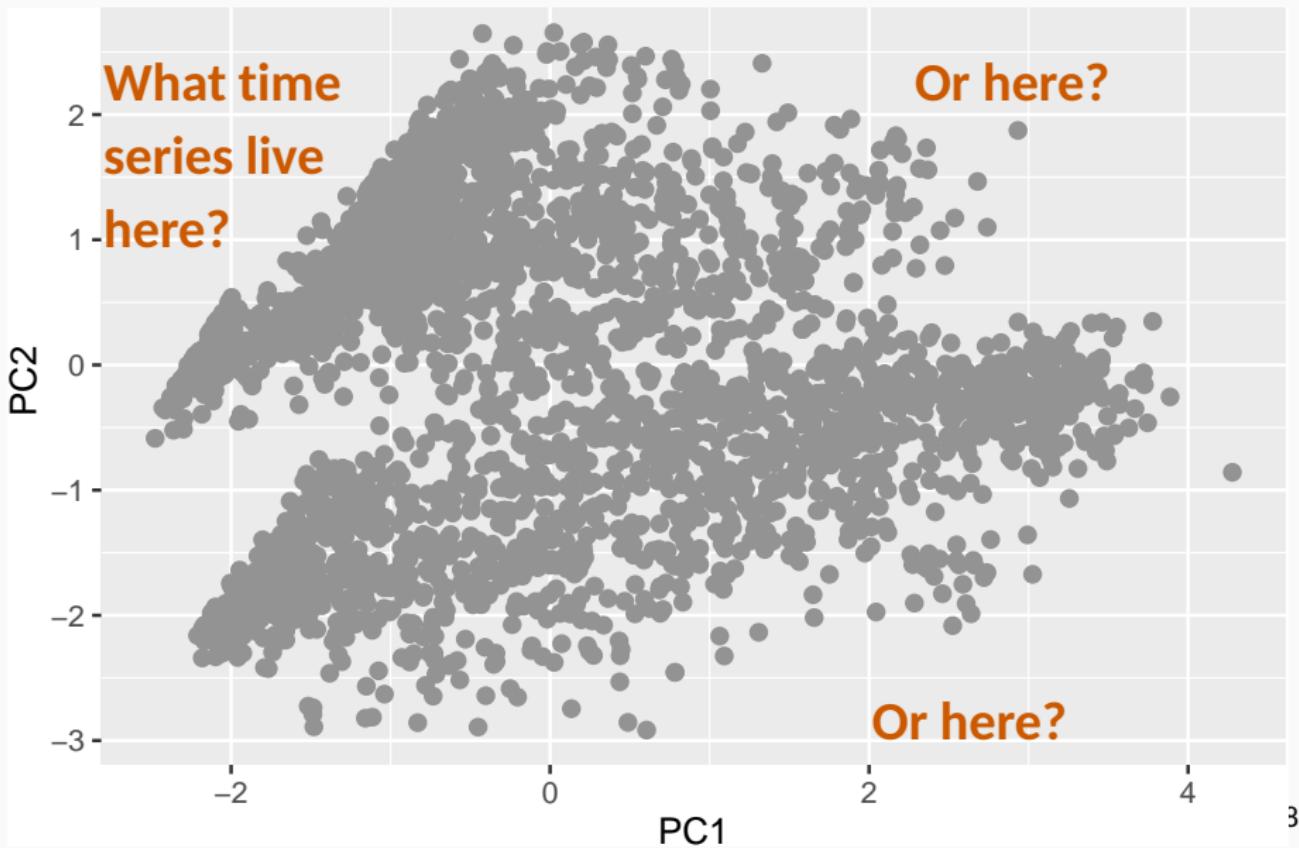
What about the holes?



What about the holes?



What about the holes?



Generating new time series

We can use the feature space to:

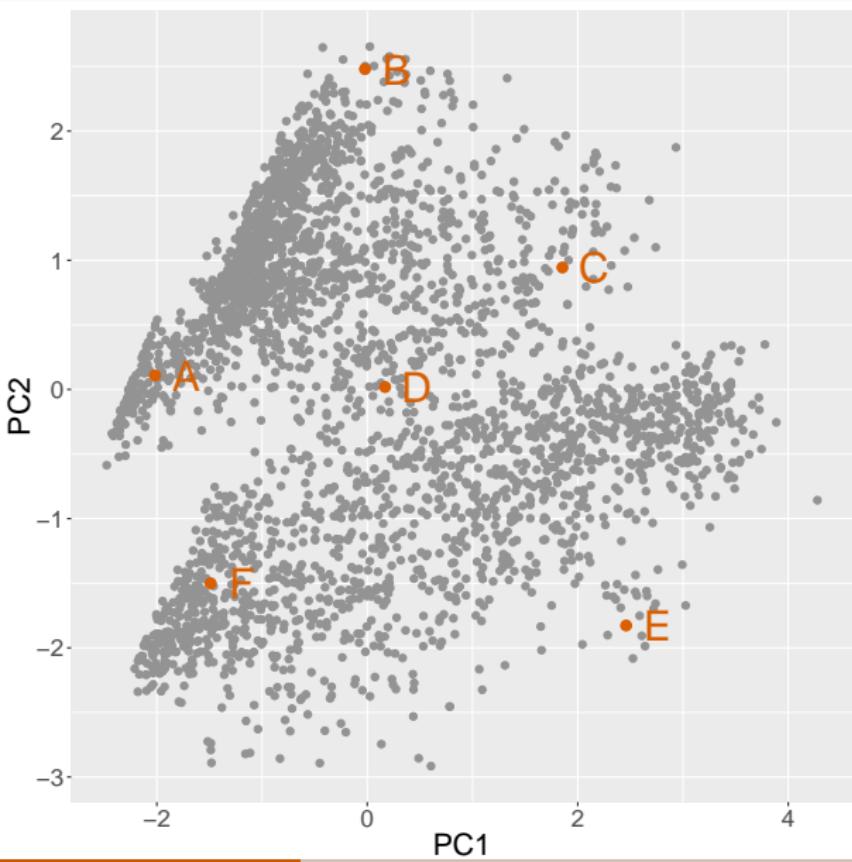
- Generate new time series with similar features to existing series
- Generate new time series where there are “holes” in the feature space.

Generating new time series

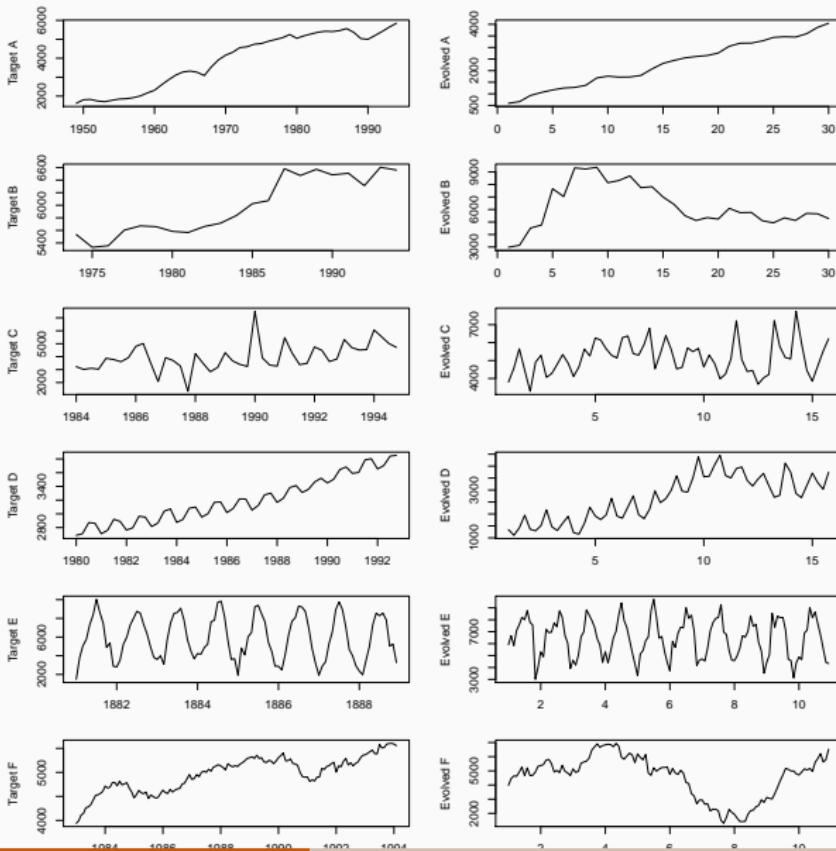
We can use the feature space to:

- Generate new time series with similar features to existing series
- Generate new time series where there are “holes” in the feature space.
- Let $\{PC_1, PC_2, \dots, PC_n\}$ be a “population” of time series of specified length and period.
- Genetic algorithm uses a process of selection, crossover and mutation to evolve the population towards a target point T_i .
- Optimize: Fitness (PC_j) = $-\sqrt{(|PC_j - T_i|^2)}$.
- Initial population random with some series in neighbourhood of T_i .

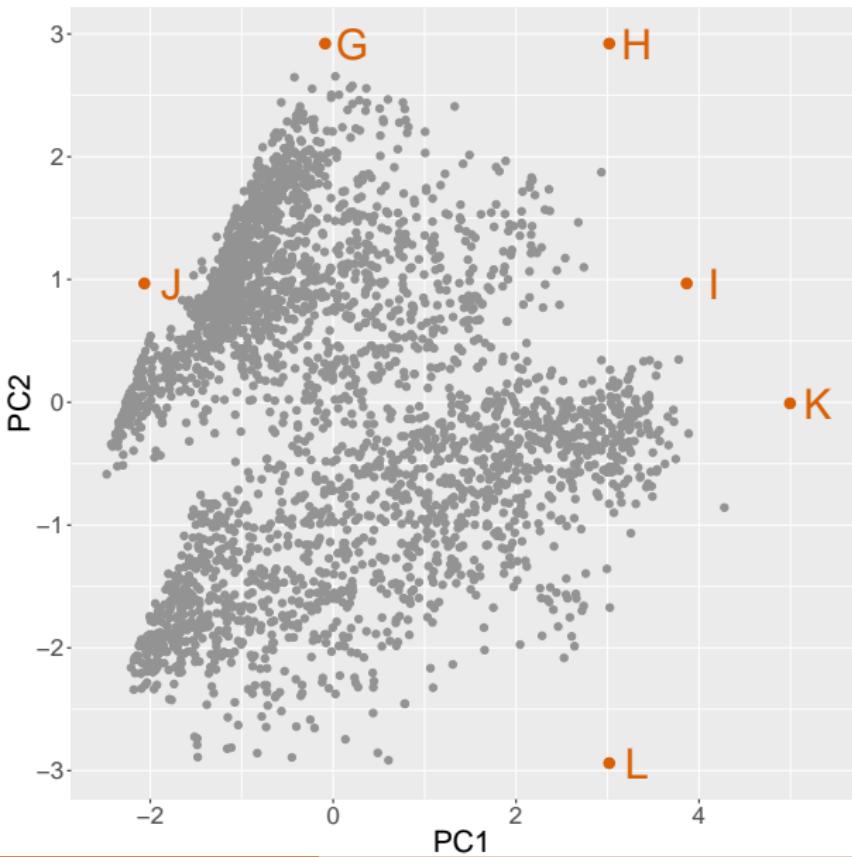
Evolving new time series



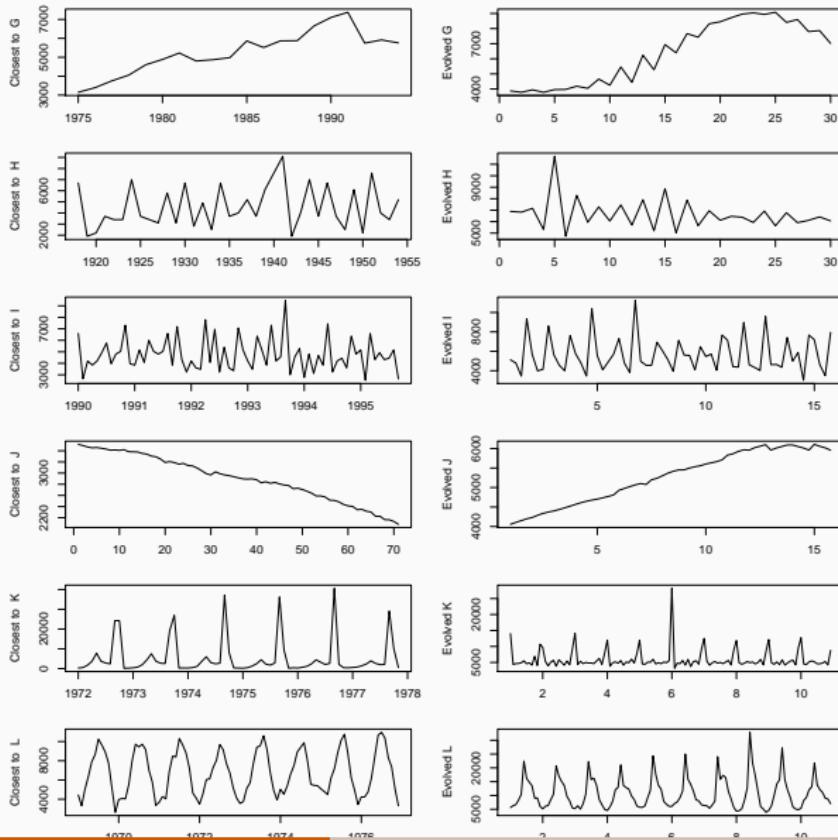
Evolving new time series



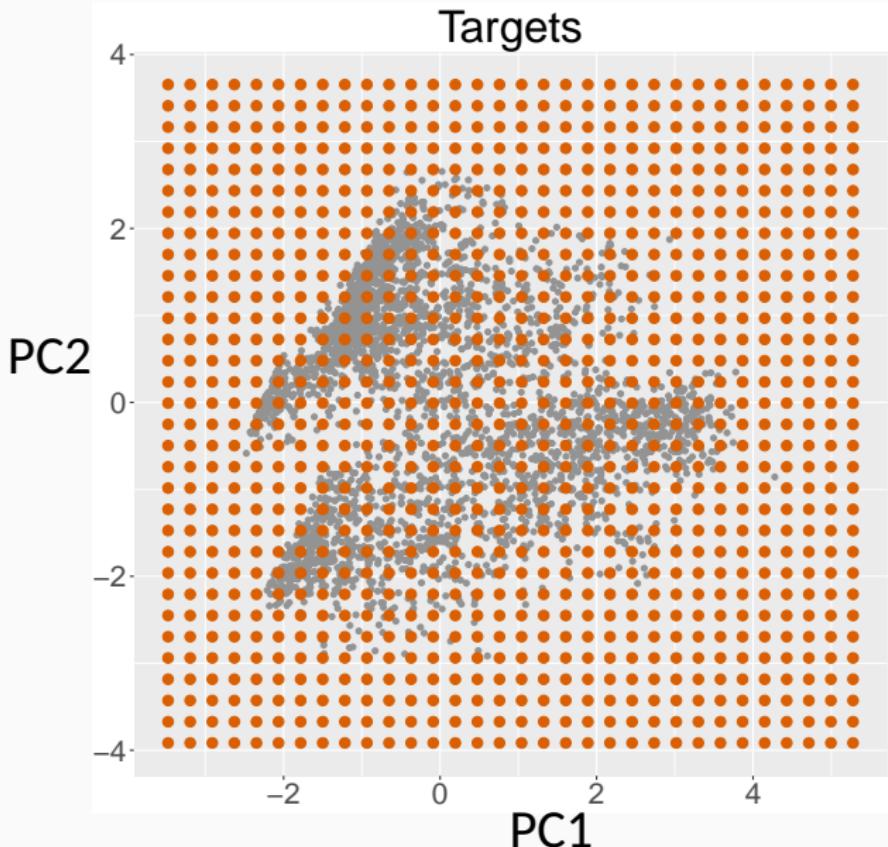
Evolving new time series



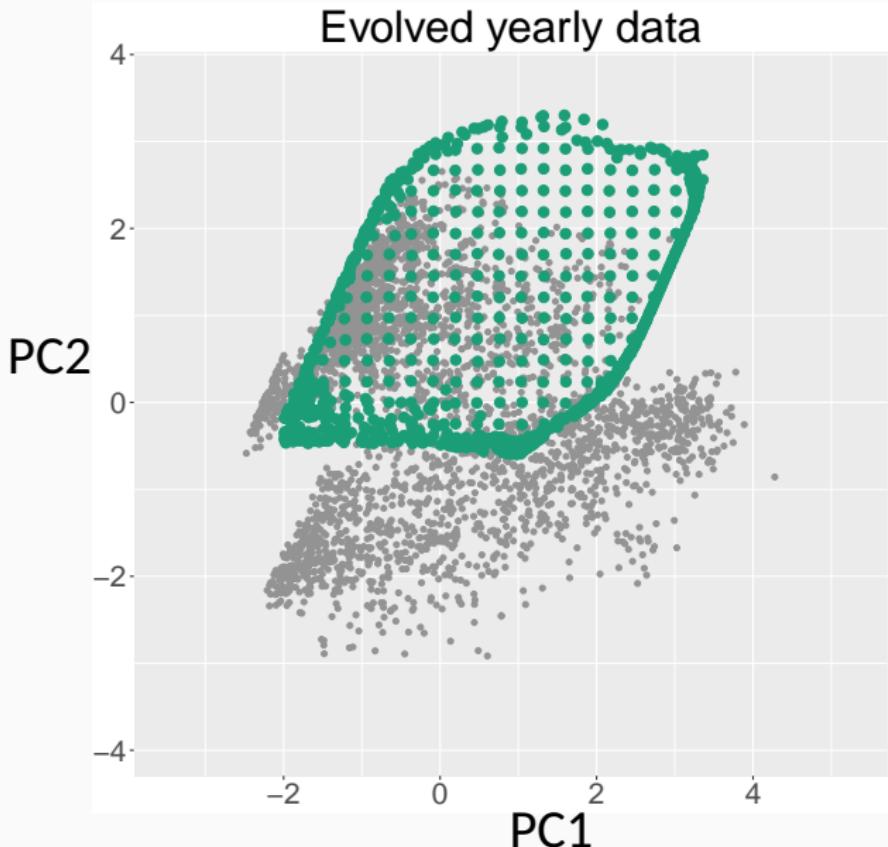
Evolving new time series



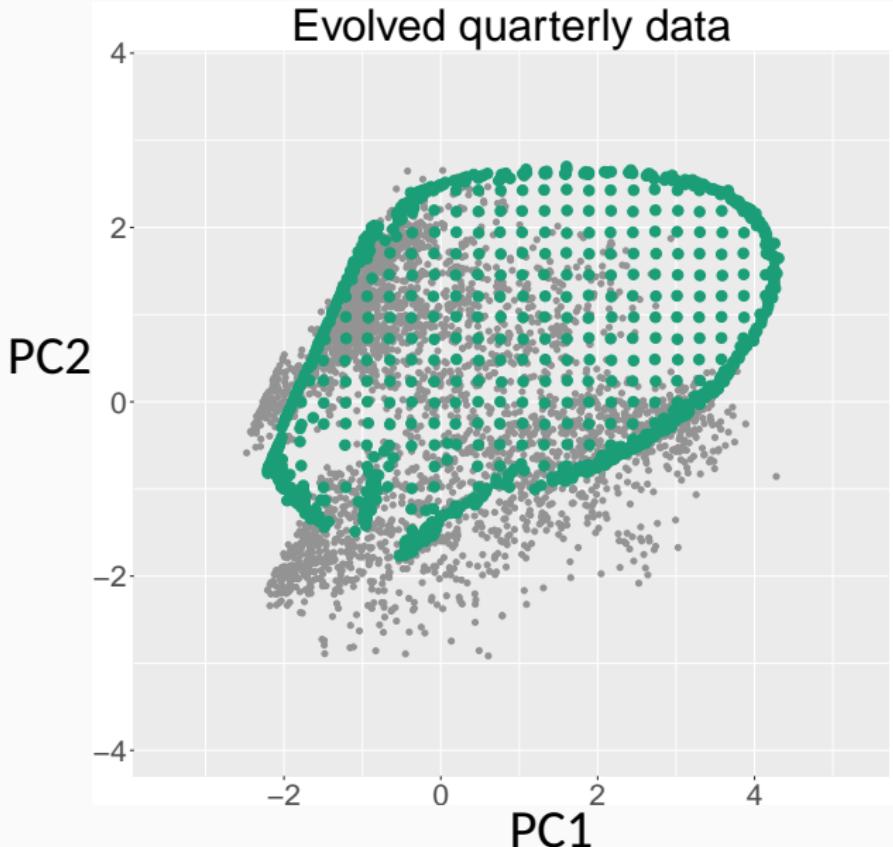
Evolving new time series



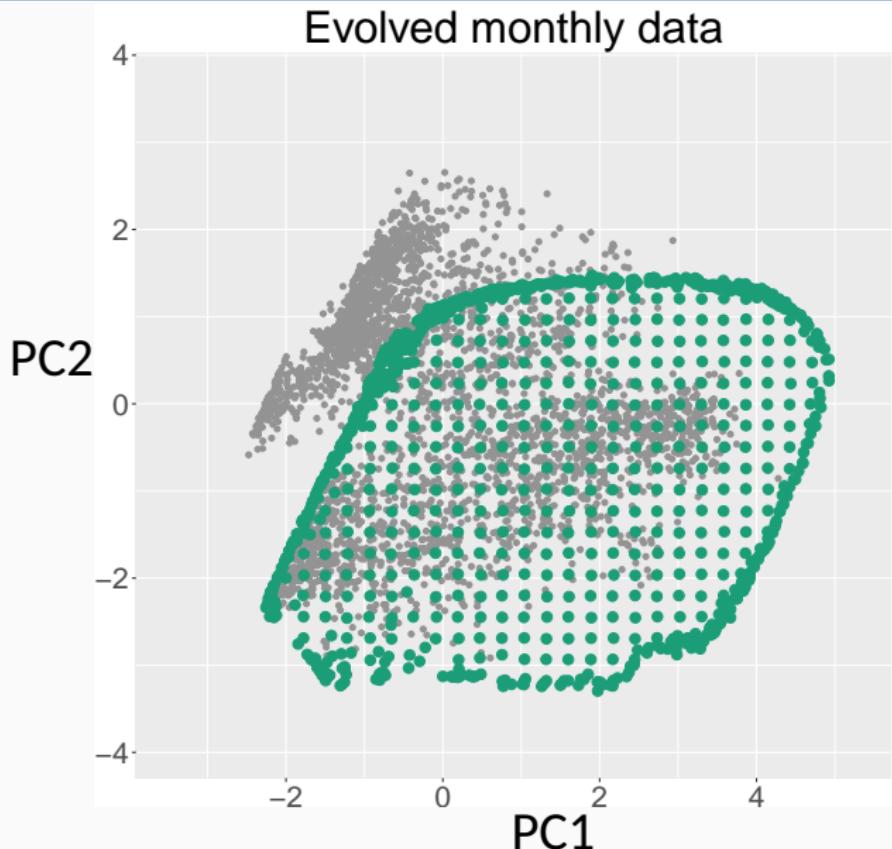
Evolving new time series



Evolving new time series



Evolving new time series



Outline

- 1 Feature-based visualization
- 2 R packages
- 3 Feature-based anomaly detection
- 4 Feature-based forecasting



Overview



Feature Extraction And Statistics for Time Series

- works with tidy temporal data provided by the `tsibble` package.
- produces time series features, decompositions, statistical summaries and visualisations.

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose  Trips
##       <qtr> <chr>    <chr> <chr>    <dbl>
## 1 1998   Q1 Adelaide SA Business  135.
## 2 1998   Q2 Adelaide SA Business  110.
## 3 1998   Q3 Adelaide SA Business  166.
## 4 1998   Q4 Adelaide SA Business  127.
## 5 1999   Q1 Adelaide SA Business  137.
## 6 1999   Q2 Adelaide SA Business  200.
## 7 1999   Q3 Adelaide SA Business  169.
## 8 1999   Q4 Adelaide SA Business  134.
## 9 2000   Q1 Adelaide SA Business  154.
## 10 2000  Q2 Adelaide SA Business  169.
## # ... with 24,310 more rows
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index    <chr>   <chr>  <chr>   <dbl>
## 1 1998   Q1 Adelaide SA  Business 135.
## 2 1998   Q2 Adelaide SA  Business 110.
## 3 1998   Q3 Adelaide SA  Business 166.
## 4 1998   Q4 Adelaide SA  Business 127.
## 5 1999   Q1 Adelaide SA  Business 137.
## 6 1999   Q2 Adelaide SA  Business 200.
## 7 1999   Q3 Adelaide SA  Business 169.
## 8 1999   Q4 Adelaide SA  Business 134.
## 9 2000   Q1 Adelaide SA  Business 154.
## 10 2000  Q2 Adelaide SA  Business 169.
## # ... with 24,310 more rows
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index    Keys          <dbl>
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
## # ... with 24,310 more rows
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index    Keys          Measure
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
## # ... with 24,310 more rows
```

tsibble objects

tourism

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:      Region, State, Purpose [304]
##   Quarter Region  State Purpose Trips
##   Index    Keys          Measure
## 1 1998 Q1 Adelaide SA Business 135.
## 2 1998 Q2 Adelaide SA Business 110.
## 3 1998 Q3 Adelaide SA Business 166.
## 4 1998 Q4 Adelaide SA Business 127.
## 5 1999 Q1 Adelaide SA Business 137.
## 6 1999 Q2 Adelaide SA Business 200.
## 7 1999 Q3 Adelaide SA Business 169.
## 8 1999 Q4 Adelaide SA Business 134.
## 9 2000 Q1 Adelaide SA Business 154.
## 10 2000 Q2 Adelaide SA Business 169.
## # ... with 24,310 more rows
```

Domestic visitor
nights in thousands
by state/region and
purpose.

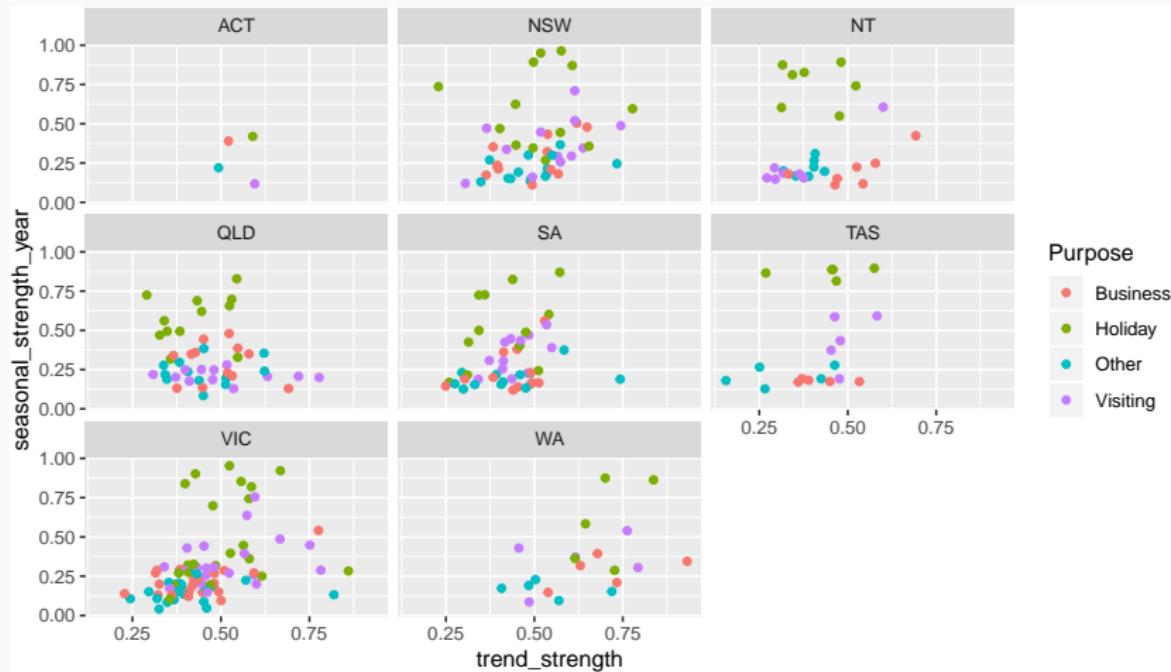
Feature extraction and statistics

```
tourism %>% features(Trips, feature_set(tags="stl"))

## # A tibble: 304 x 12
##   Region State Purpose trend_strength seasonal_streng~
##   <chr>   <chr>  <chr>          <dbl>            <dbl>
## 1 Adela~ SA    Busine~        0.451            0.380
## 2 Adela~ SA    Holiday       0.541            0.601
## 3 Adela~ SA    Other         0.743            0.189
## 4 Adela~ SA    Visiti~       0.433            0.446
## 5 Adela~ SA    Busine~        0.453            0.140
## 6 Adela~ SA    Holiday       0.512            0.244
## 7 Adela~ SA    Other         0.584            0.374
## 8 Adela~ SA    Visiti~       0.481            0.228
## 9 Alice~ NT    Busine~        0.526            0.224
## 10 Alice~ NT   Holiday       0.377            0.827
## # ... with 294 more rows, and 7 more variables:
## #   seasonal_peak_year <dbl>, seasonal_trough_year <dbl>,
## #   spikiness <dbl>, linearity <dbl>, curvature <dbl>,
## #   stl_e_acf1 <dbl>, stl_e_acf10 <dbl>
```

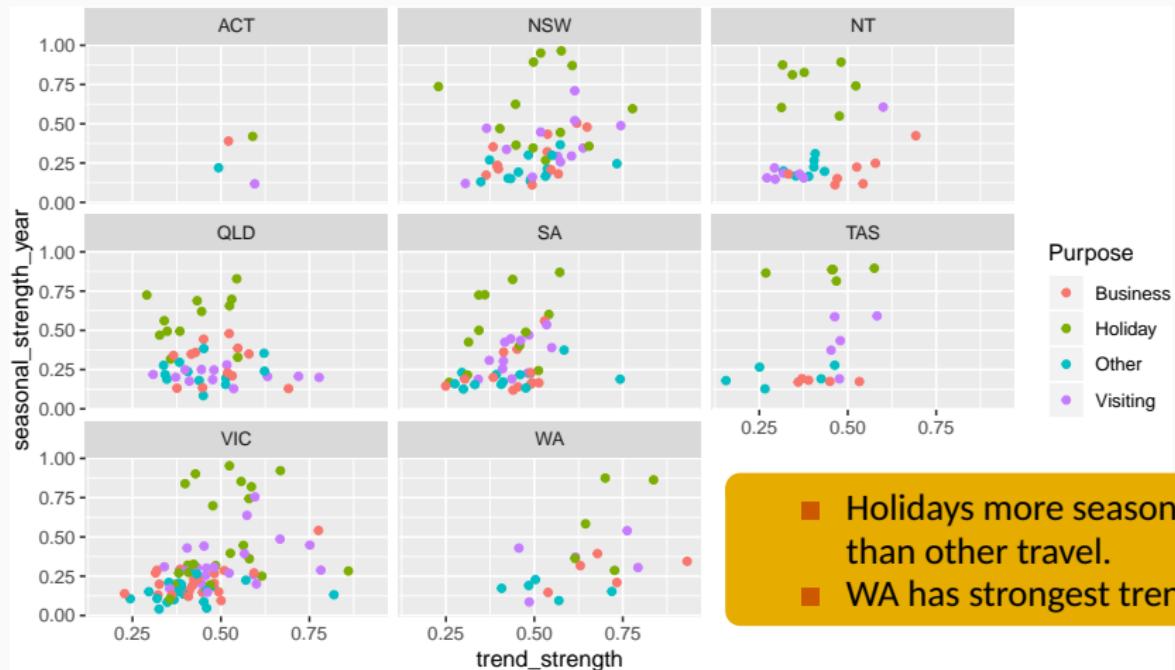
Feature extraction and statistics

```
tourism %>% features(Trips, feature_set(tags="stl")) %>%
  ggplot(aes(x=trend_strength, y=seasonal_strength_year, col=Purpose)) +
  geom_point() + facet_wrap(vars(State))
```



Feature extraction and statistics

```
tourism %>% features(Trips, feature_set(tags="stl")) %>%
  ggplot(aes(x=trend_strength, y=seasonal_strength_year, col=Purpose)) +
  geom_point() + facet_wrap(vars(State))
```



Feature extraction and statistics

Find the most seasonal time series:

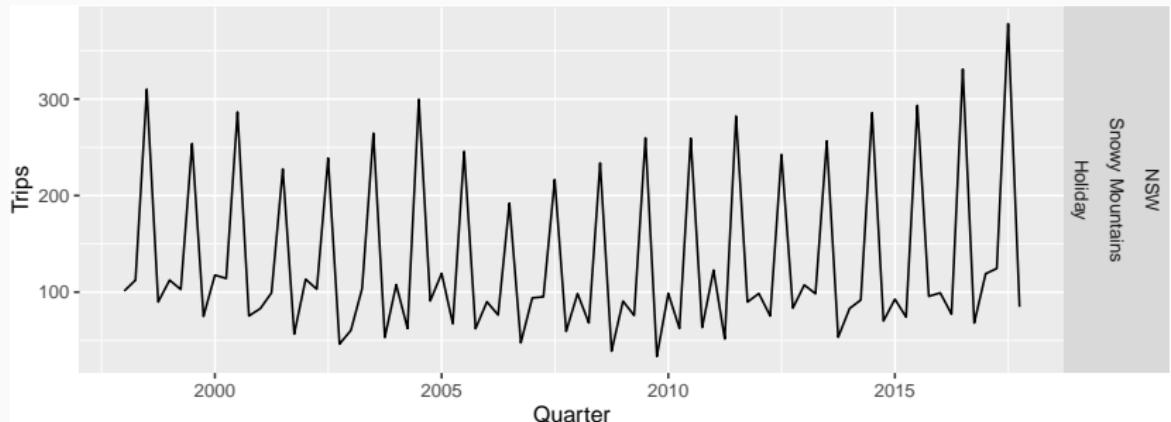
```
most_seasonal <- tourism %>%
  features(Trips, feature_set(tags="stl")) %>%
  filter(seasonal_strength_year == max(seasonal_strength_year))
```

Feature extraction and statistics

Find the most seasonal time series:

```
most_seasonal <- tourism %>%
  features(Trips, feature_set(tags="stl")) %>%
  filter(seasonal_strength_year == max(seasonal_strength_year))

tourism %>%
  right_join(most_seasonal, by = c("State","Region","Purpose")) %>%
  ggplot(aes(x = Quarter, y = Trips)) + geom_line() +
  facet_grid(vars(State,Region,Purpose))
```



Feature extraction and statistics

```
tourism_features <- tourism %>%  
  features(Trips, feature_set(pkgs="feasts"))  
  
## # A tibble: 304 x 47  
##   Region State Purpose trend_strength seasonal_streng~  
##   <chr>   <chr>  <chr>           <dbl>            <dbl>  
## 1 Adela~  SA     Busine~        0.451            0.380  
## 2 Adela~  SA     Holiday        0.541            0.601  
## 3 Adela~  SA     Other          0.743            0.189  
## 4 Adela~  SA     Visiti~       0.433            0.446  
## 5 Adela~  SA     Busine~        0.453            0.140  
## 6 Adela~  SA     Holiday        0.512            0.244  
## 7 Adela~  SA     Other          0.584            0.374  
## 8 Adela~  SA     Visiti~       0.481            0.228  
## 9 Alice~   NT     Busine~        0.526            0.224  
## 10 Alice~  NT    Holiday        0.377            0.827  
## # ... with 294 more rows, and 42 more variables:  
## #   seasonal_peak_year <dbl>, seasonal_trough_year <dbl>,  
## #   spikiness <dbl>, linearity <dbl>, curvature <dbl>,  
## #   stl_e_acf1 <dbl>, stl_e_acf10 <dbl>, acf1 <dbl>,  
## #   acf10 <dbl>, diff1_acf1 <dbl>, diff1_acf10 <dbl>,  
## #   ...
```

All features from
the feasts
package

Feature extraction and statistics

```
pcs <- tourism_features %>% select(-State, -Region, -Purpose) %>%  
  prcomp(scale=TRUE) %>% augment(tourism_features)
```

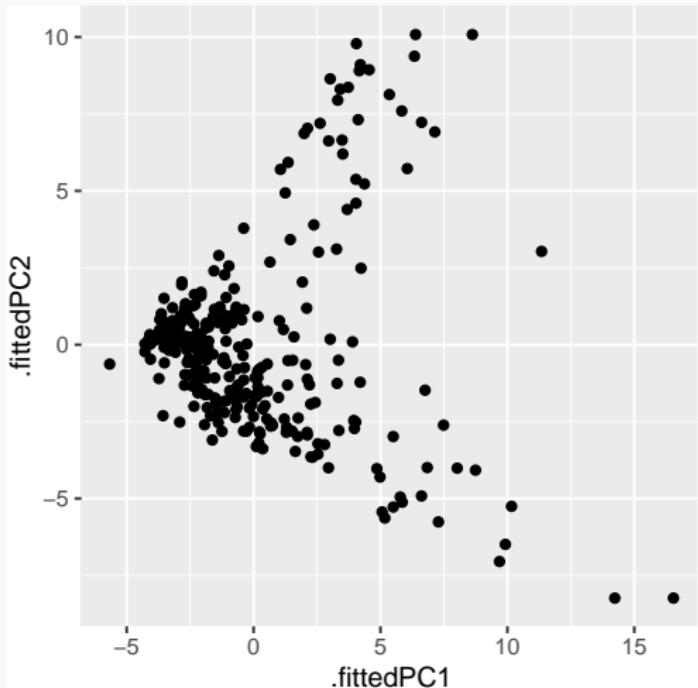
```
## # A tibble: 304 x 92  
##   .rownames Region State Purpose trend_strength  
##   <fct>     <chr>  <chr>  <chr>          <dbl>  
## 1 1          Adela~ SA     Busine~        0.451  
## 2 2          Adela~ SA     Holiday         0.541  
## 3 3          Adela~ SA     Other           0.743  
## 4 4          Adela~ SA     Visiti~        0.433  
## 5 5          Adela~ SA     Busine~        0.453  
## 6 6          Adela~ SA     Holiday         0.512  
## 7 7          Adela~ SA     Other           0.584  
## 8 8          Adela~ SA     Visiti~        0.481  
## 9 9          Alice~ NT    Busine~        0.526  
## 10 10        Alice~ NT    Holiday         0.377  
## # ... with 294 more rows, and 87 more variables:  
## #   seasonal_strength_year <dbl>, seasonal_peak_year <dbl>,  
## #   seasonal_trough_year <dbl>, spikiness <dbl>,  
## #   linearity <dbl>, curvature <dbl>, stl_e_acf1 <dbl>,  
## #   stl_e_acf10 <dbl>, acf1 <dbl>, acf10 <dbl>,  
## #   liss1 ~ f1 ~ f2 ~ liss1 ~ f12 ~ f22 ~ liss2 ~ f1 ~ f2
```

Principal components based on all features from the feasts package

Feature extraction and statistics

```
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2)) +  
  geom_point() + theme(aspect.ratio=1)
```

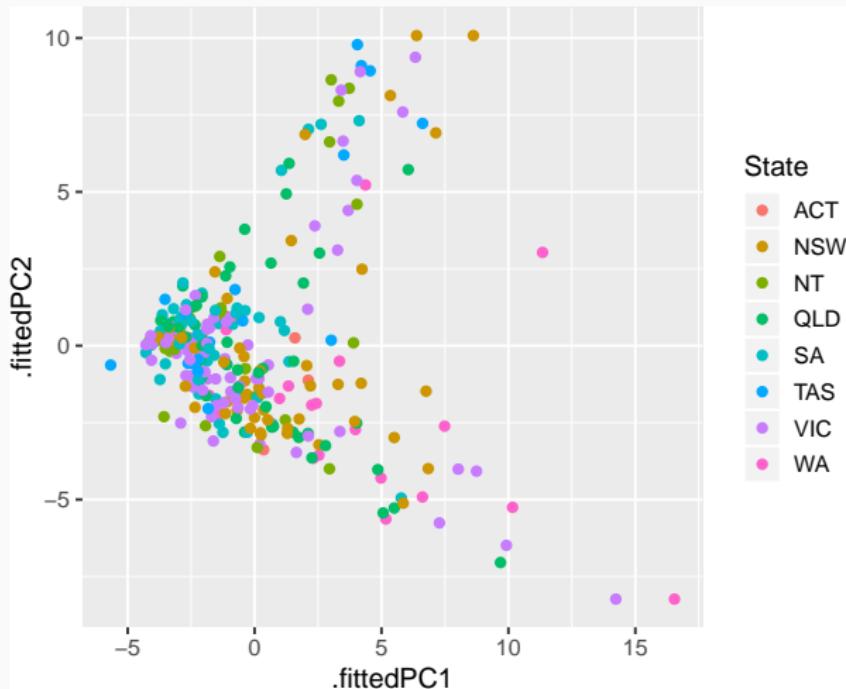
Principal components
based on all features
from the feasts
package



Feature extraction and statistics

```
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2, col=State)) +  
  geom_point() + theme(aspect.ratio=1)
```

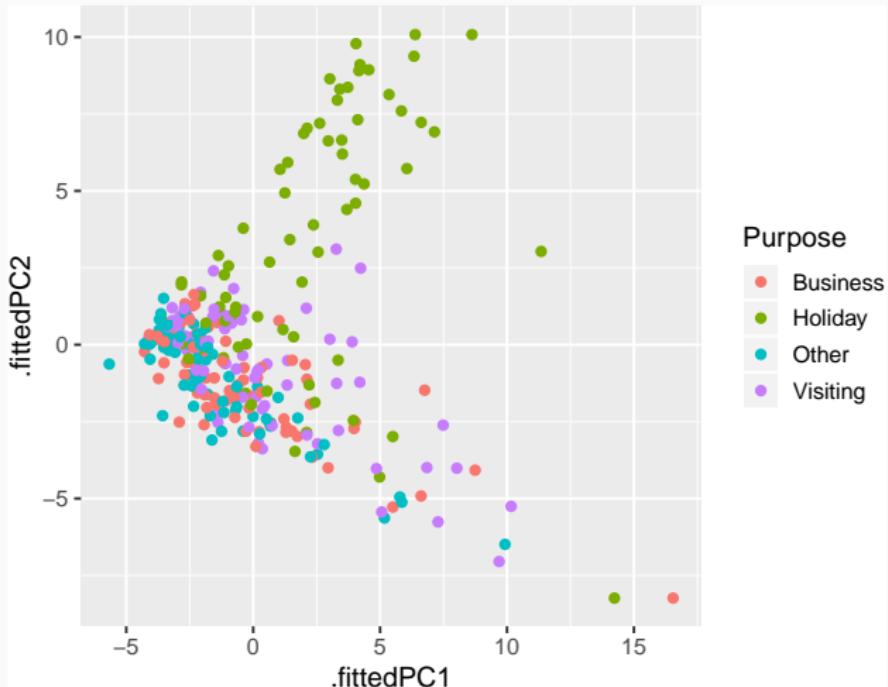
Principal components
based on all features
from the feasts
package



Feature extraction and statistics

```
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2, col=Purpose)) +  
  geom_point() + theme(aspect.ratio=1)
```

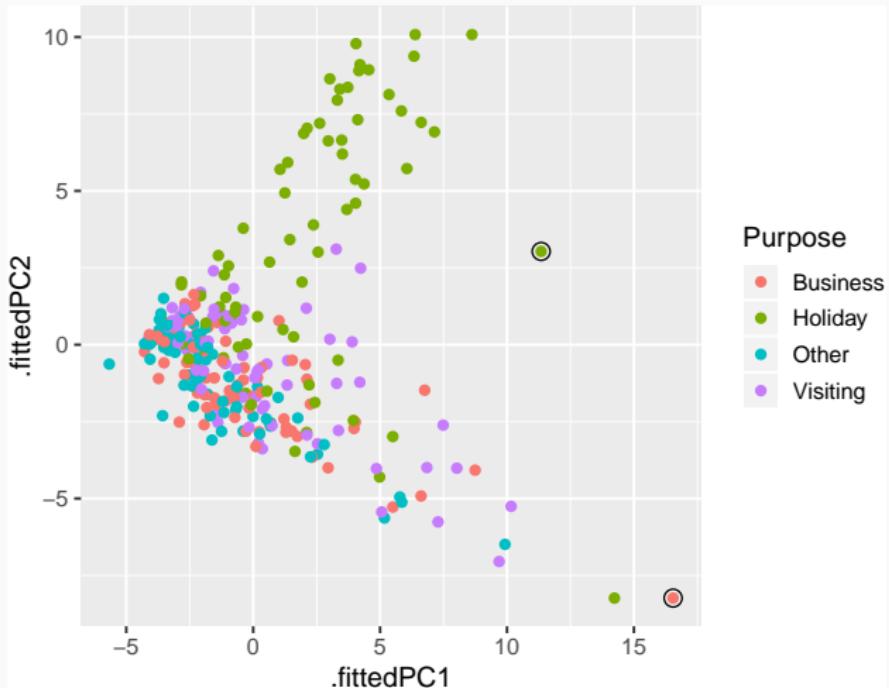
Principal components
based on all features
from the feasts
package



Feature extraction and statistics

```
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2, col=Purpose)) +  
  geom_point() + theme(aspect.ratio=1)
```

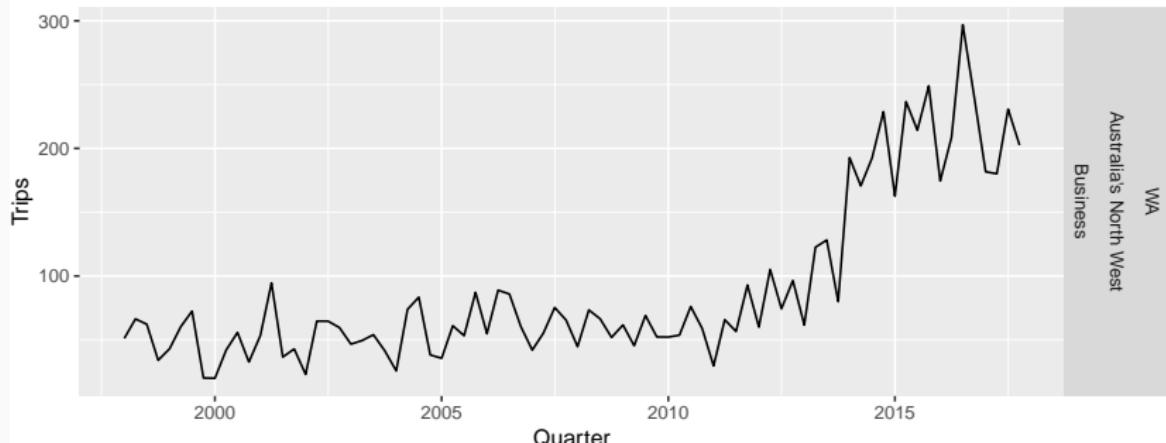
Principal components
based on all features
from the feasts
package



Feature extraction and statistics

```
pcs %>% filter(.fittedPC1 == max(.fittedPC1)) %>%
  left_join(tourism, by = c("State", "Region", "Purpose")) %>%
  ggplot(aes(x = Quarter, y = Trips)) +
  geom_line() +
  facet_grid(vars(State, Region, Purpose)) +
  ggtitle("Outlying time series in PC space") +
  theme(legend.position = "none")
```

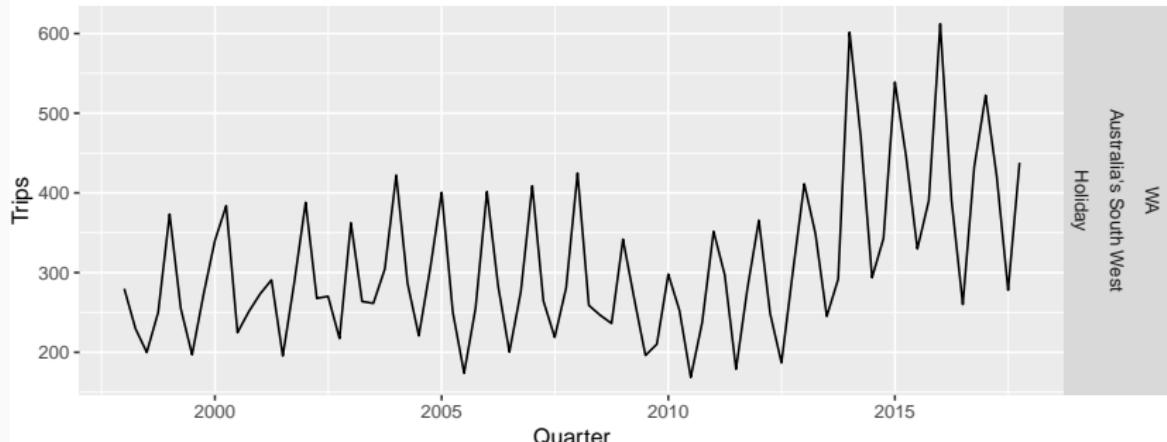
Outlying time series in PC space



Feature extraction and statistics

```
pcs %>% filter(.fittedPC1 > 10 & .fittedPC2 > 2.5) %>%  
  left_join(tourism, by = c("State", "Region", "Purpose")) %>%  
  ggplot(aes(x = Quarter, y = Trips)) +  
    geom_line() +  
    facet_grid(vars(State, Region, Purpose)) +  
    ggtitle("Outlying time series in PC space") +  
    theme(legend.position = "none")
```

Outlying time series in PC space

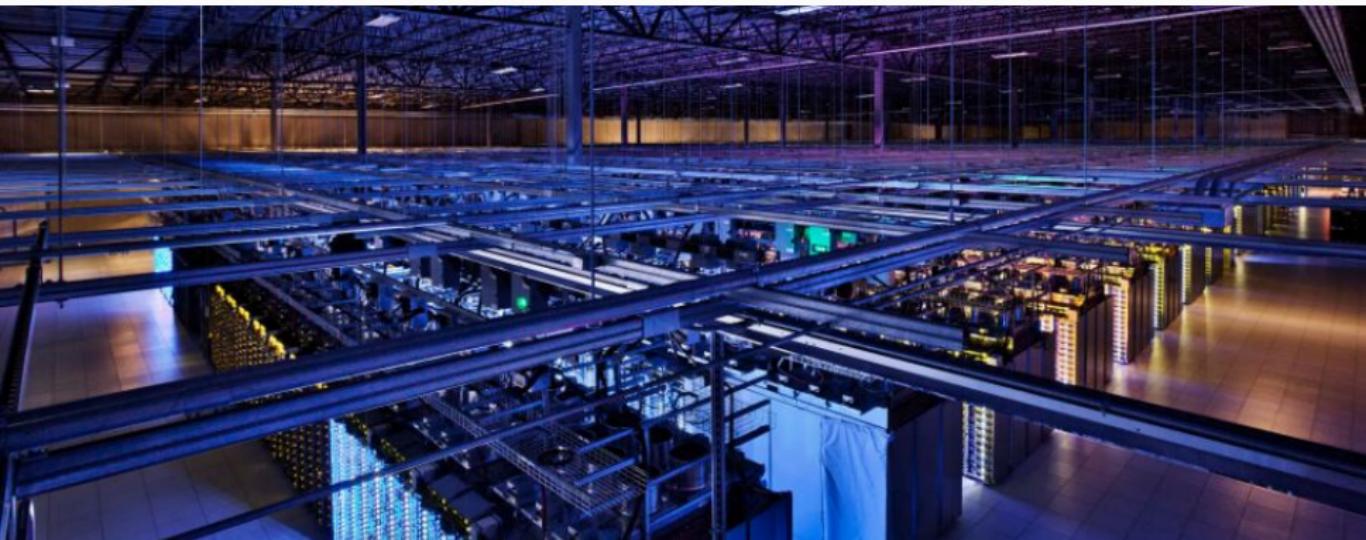


Outline

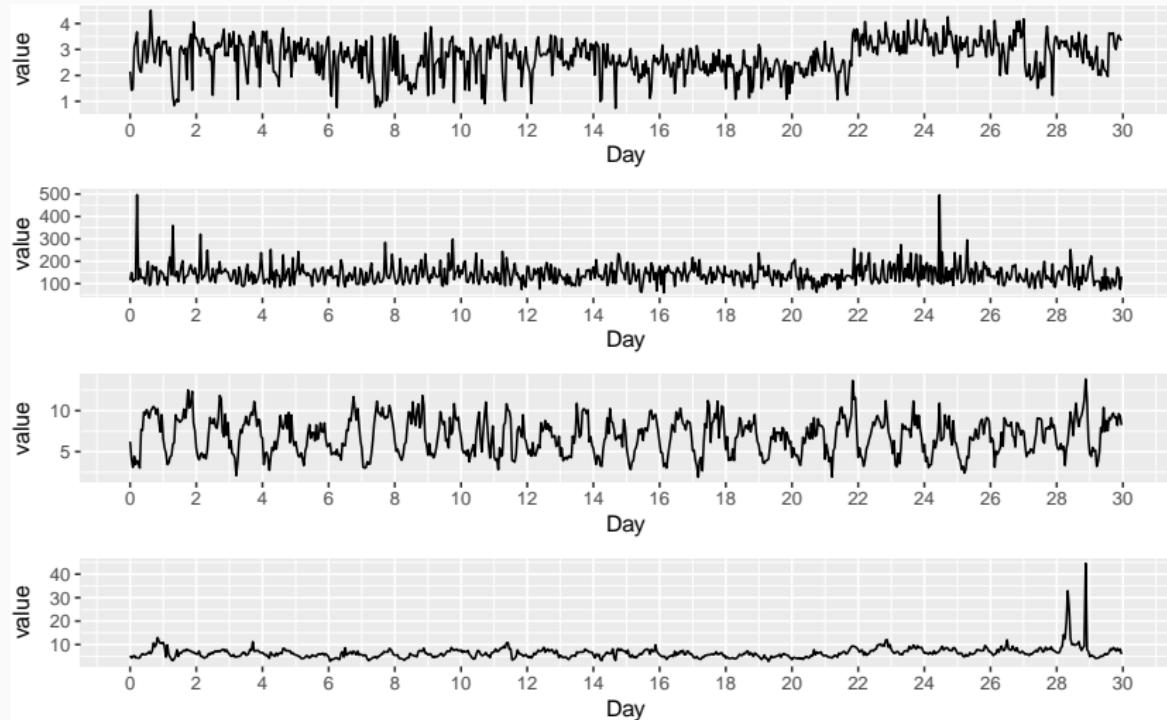
- 1 Feature-based visualization
- 2 R packages
- 3 Feature-based anomaly detection
- 4 Feature-based forecasting

Yahoo server metrics

- Tens of thousands of time series collected at one-hour intervals over 1–2 months.
- Consisting of several server metrics (e.g. CPU usage and paging views) from many server farms globally.
- Aim: find unusual (anomalous) time series.



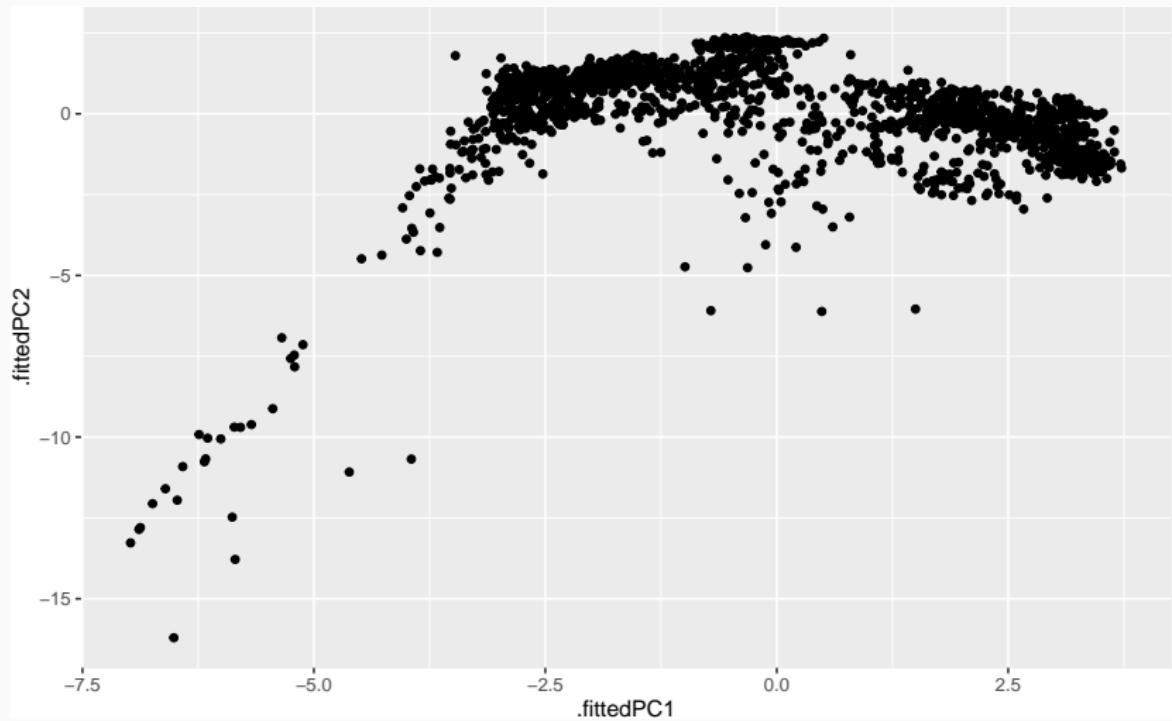
Yahoo server metrics



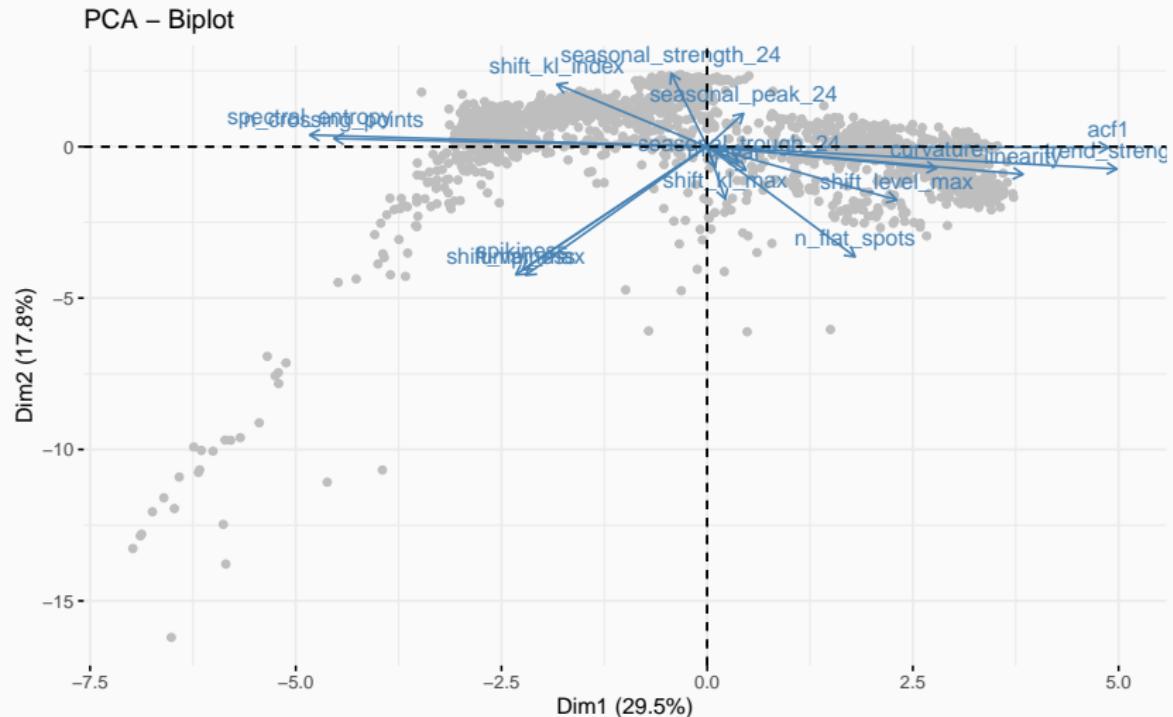
Yahoo server metrics

- **ACF1:** first order autocorrelation = $\text{Corr}(Y_t, Y_{t-1})$
- Strength of **trend** and **seasonality** based on STL
- 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

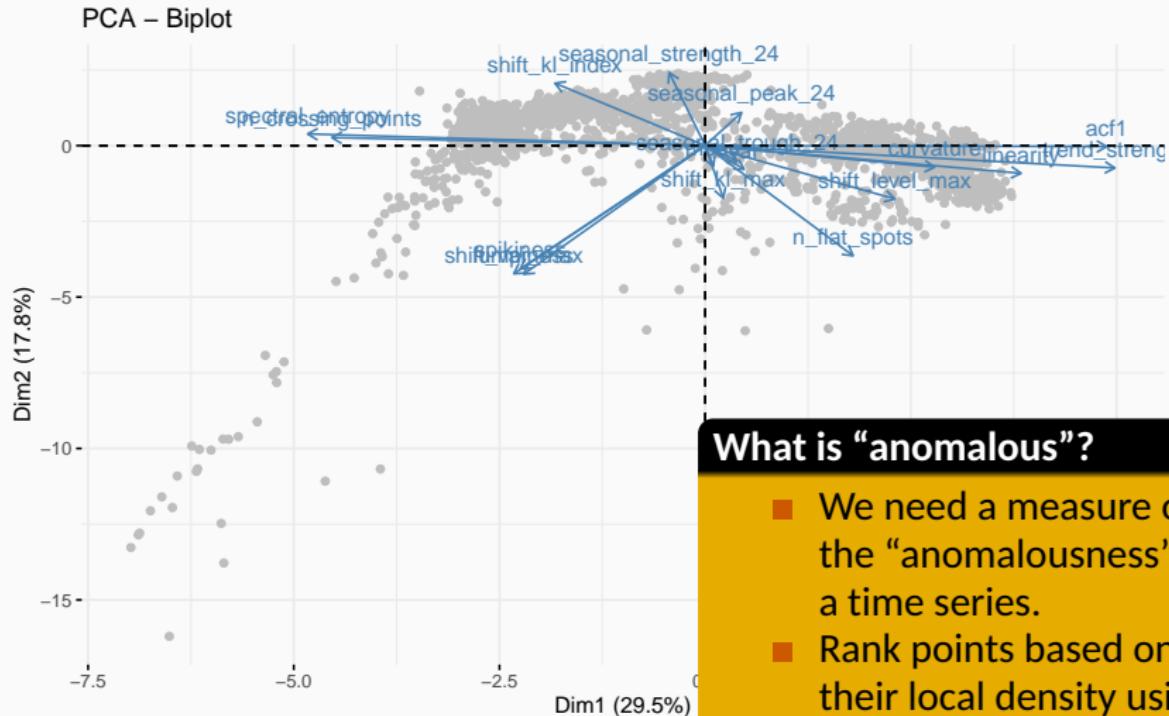
Feature space



Feature space



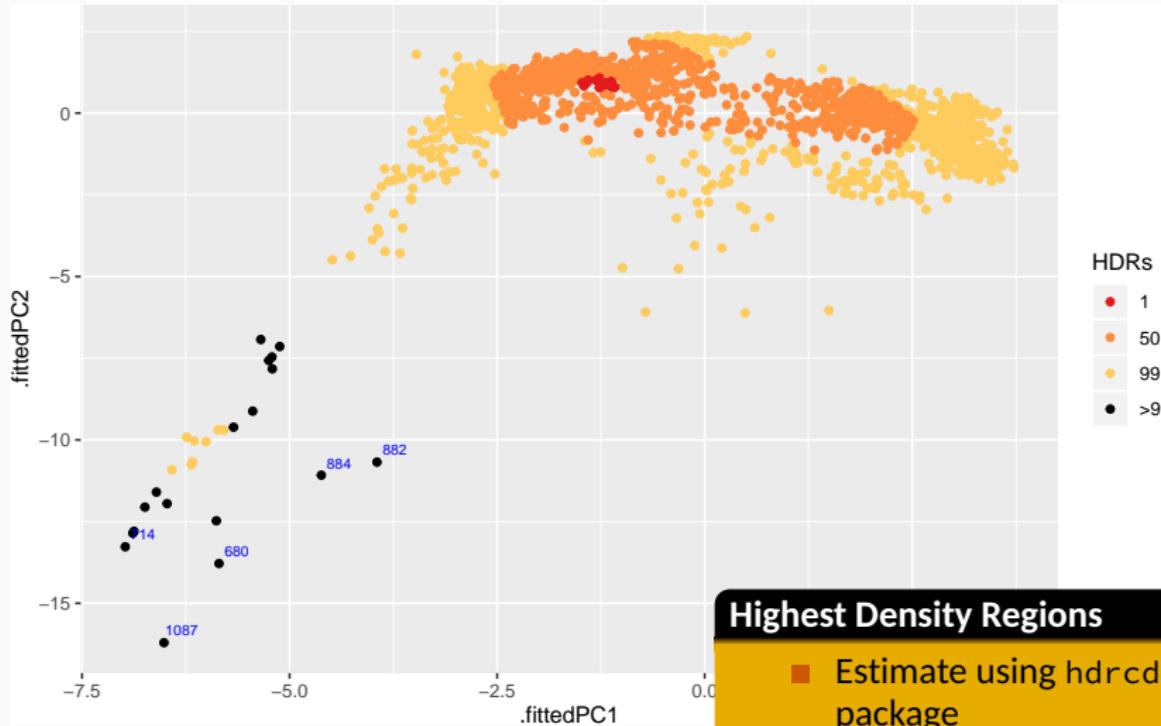
Feature space



What is “anomalous”?

- We need a measure of the “anomalousness” of a time series.
- Rank points based on their local density using a bivariate kernel density estimate.

Finding weird time series



Highest Density Regions

- Estimate using `hdr.cde` package
- Highlight outlying points as those with lowest density.

Outline

- 1 Feature-based visualization
- 2 R packages
- 3 Feature-based anomaly detection
- 4 Feature-based forecasting

M competition: 1982

Journal of Forecasting, Vol. 1, 111–153 (1982)

The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition

S. MAKRIDAKIS
INSEAD, Fontainebleau, France

A. ANDERSEN
University of Sydney, Australia

R. CARBONE
Université Laval, Quebec, Canada

R. FILDES
Manchester Business School, Manchester, England

M. HIBON
INSEAD, Fontainebleau, France

R. LEWANDOWSKI
Marketing Systems, Essen, Germany

J. NEWTON
E. PARZEN
Texas A & M University, Texas, U.S.A.

R. WINKLER
Indiana University, Bloomington, U.S.A.

ABSTRACT

In the last few decades many methods have become available for forecasting. As always, when alternatives exist, choices need to be made so that an appropriate forecasting method can be selected and used for the specific situation being considered. This paper reports the results of a forecasting competition that provides information to facilitate such choice. Seven experts in each of the 24 methods forecasted up to 1001 series for six up to eighteen time horizons. The results of the competition are presented in this paper whose purpose is to provide empirical evidence about differences found to exist among the various extrapolative (time series) methods used in the competition.

M-competition

- 1001 series from demography, industry, economics.
- Annual, quarterly, monthly data.
- Anyone could submit forecasts.
- Multiple forecast measures used.

M3 competition: 2000



ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

www.elsevier.com/locate/ijforecast

*international journal
of forecasting*

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. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting accuracy **56**

M4 competition: 2018

International Journal of Forecasting 34 (2018) 802–808



Contents lists available at [ScienceDirect](#)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast



The M4 Competition: Results, findings, conclusion and way forward



Spyros Makridakis^{a,b,*}, Evangelos Spiliotis^c, Vassilios Assimakopoulos^c

^a University of Nicosia, Nicosia, Cyprus

^b Institute For the Future (IFF), Nicosia, Cyprus

^c Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, Greece

ARTICLE INFO

Keywords:

Forecasting competitions

M Competitions

ABSTRACT

The M4 competition is the continuation of three previous competitions started more than 45 years ago whose purpose was to learn how to improve forecasting accuracy, and

M4 competition: 2018

- January – May 2018
- 100,000 time series: yearly, quarterly, monthly, weekly, daily, hourly.
- Point forecast and prediction intervals assessed.
- Code must be public
- 248 registrations, 50 submissions.

M4 competition: 2018

- January – May 2018
- 100,000 time series: yearly, quarterly, monthly, weekly, daily, hourly.
- Point forecast and prediction intervals assessed.
- Code must be public
- 248 registrations, 50 submissions.

Winning methods

- 1 Hybrid of Recurrent Neural Network and Exponential Smoothing models
- 2 FFORMA: Feature-based forecast combinations using xgboost to find weights

Features used to select a forecasting model

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive method
- unit root test statistics
- crossing points, flat spots
- peaks, troughs
- ACF and PACF based features - calculated on raw, differenced, and remainder series.
- ARCH/GARCH statistics and ACF of squared series and residuals.

Features used to select a forecasting model

Why these features?

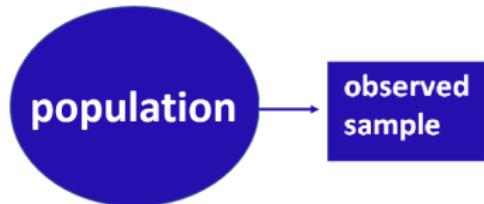
- Hyndman, Wang and Laptev. “Large scale unusual time series detection” (ICDM 2015).
- Kang, Hyndman & Smith-Miles. “Visualising forecasting algorithm performance using time series instance spaces” (IJF 2017).
- Talagala, Hyndman and Athanasopoulos. “Meta-learning how to forecast time series” (2018).
- Implemented in the **feasts** R package

FFORMS: Feature-based FORcast Model Selection

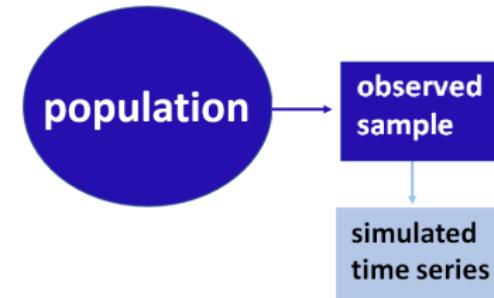


population

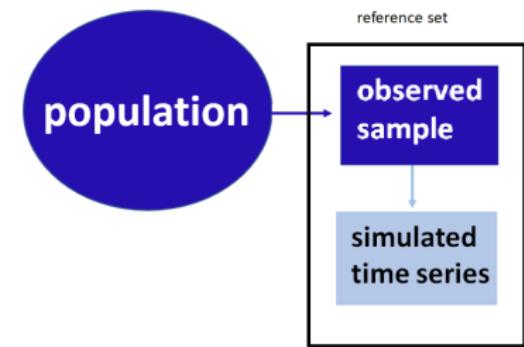
FFORMS: Feature-based FORcast Model Selection



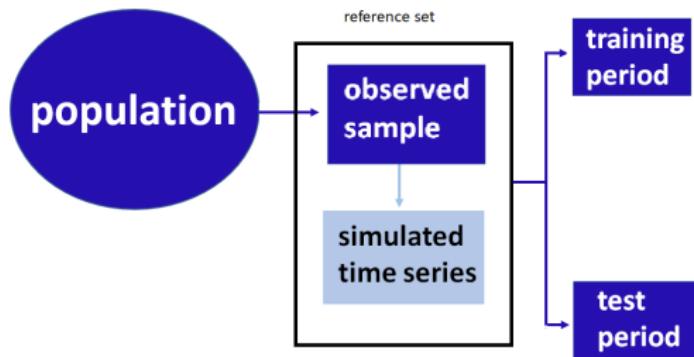
FFORMS: Feature-based FORcast Model Selection



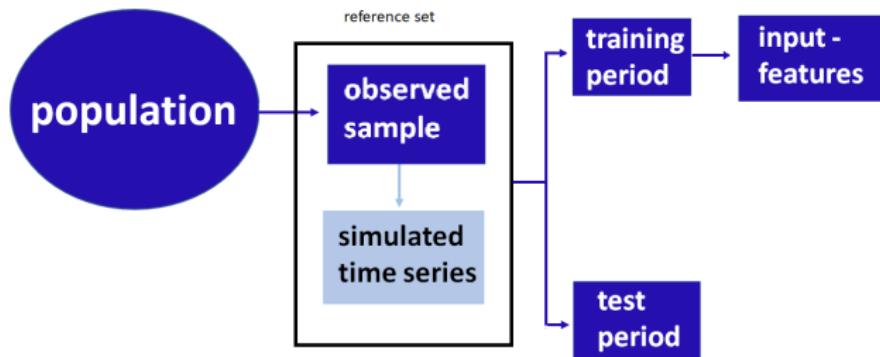
FFORMS: Feature-based FORcast Model Selection



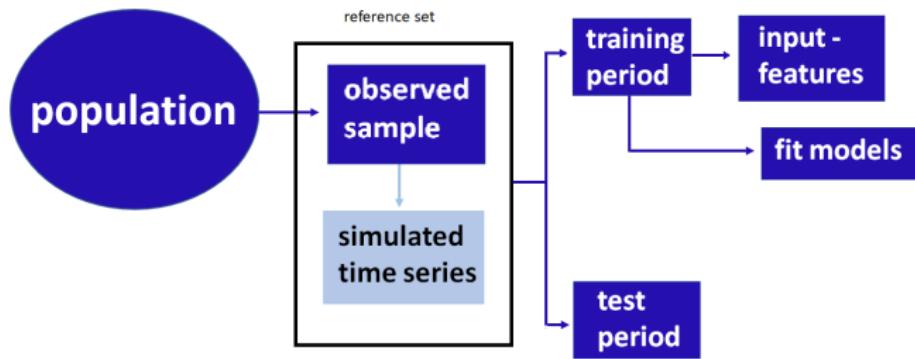
FFORMS: Feature-based FORcast Model Selection



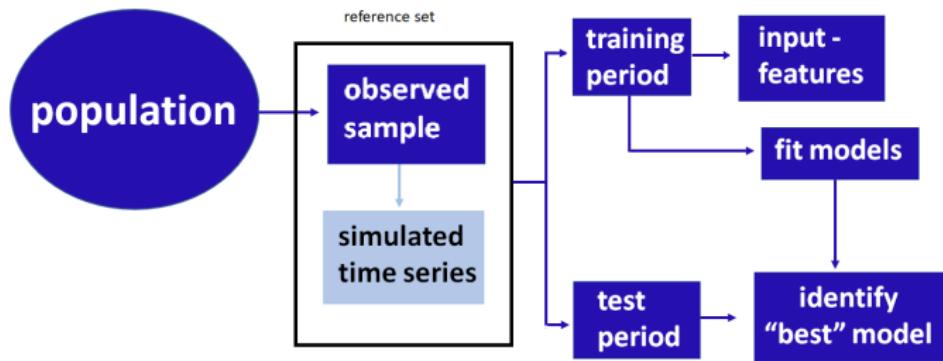
FFORMS: Feature-based FORcast Model Selection



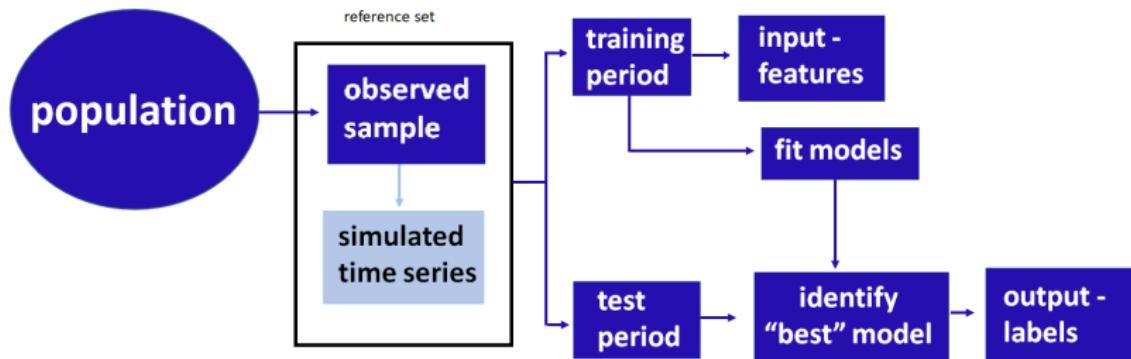
FFORMS: Feature-based FORecast Model Selection



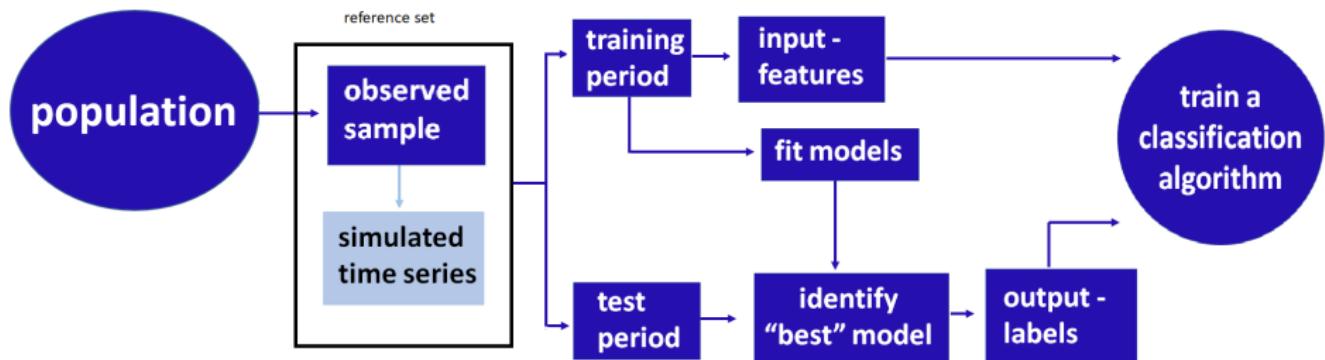
FFORMS: Feature-based FORcast Model Selection



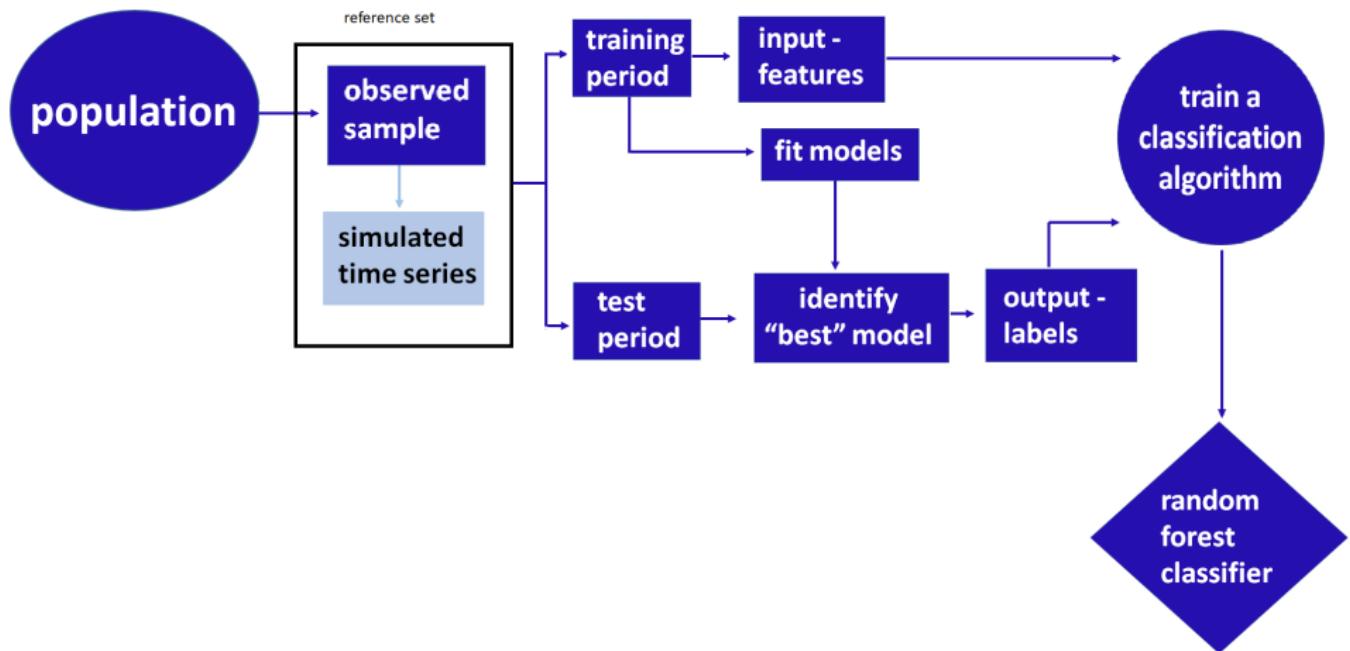
FFORMS: Feature-based FORecast Model Selection



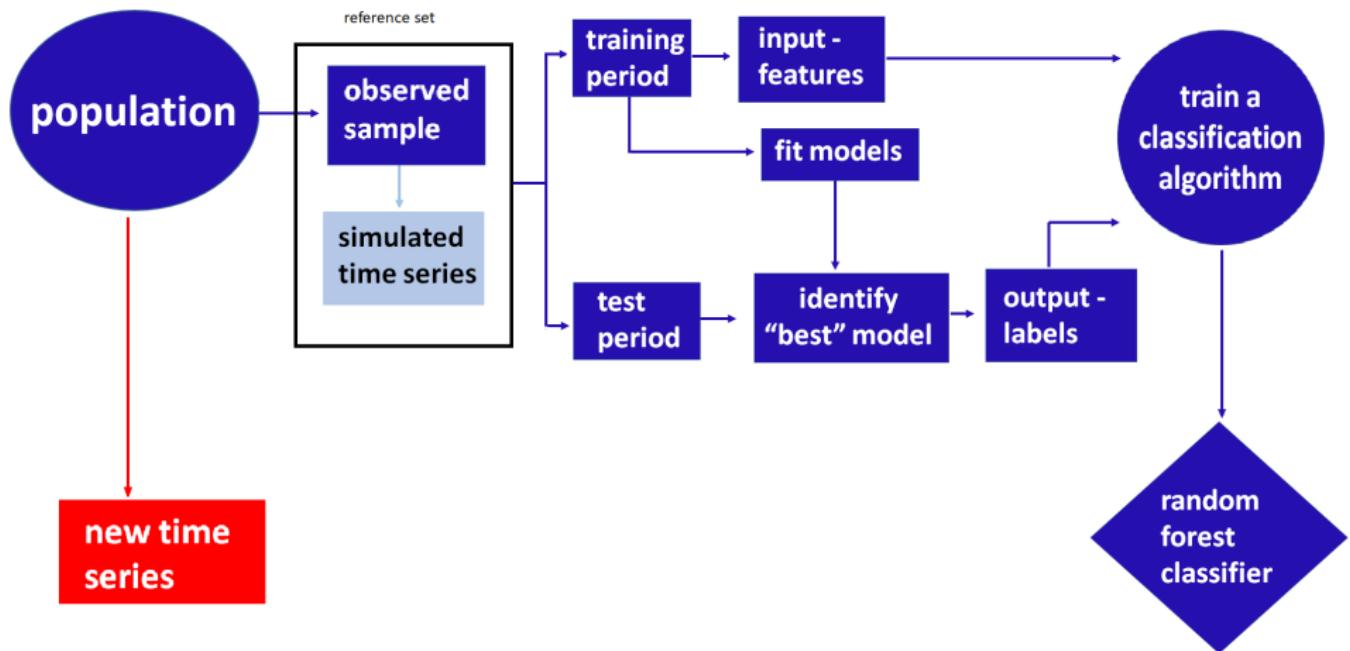
FFORMS: Feature-based FORecast Model Selection



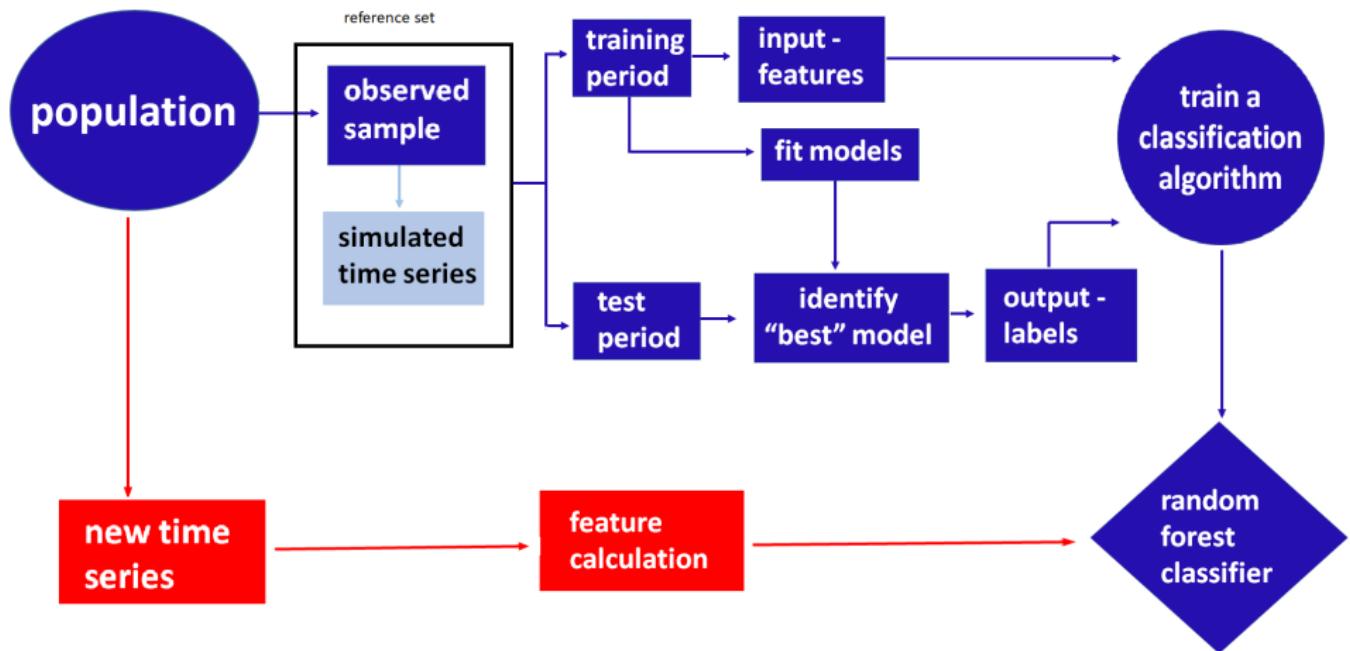
FFORMS: Feature-based FORecast Model Selection



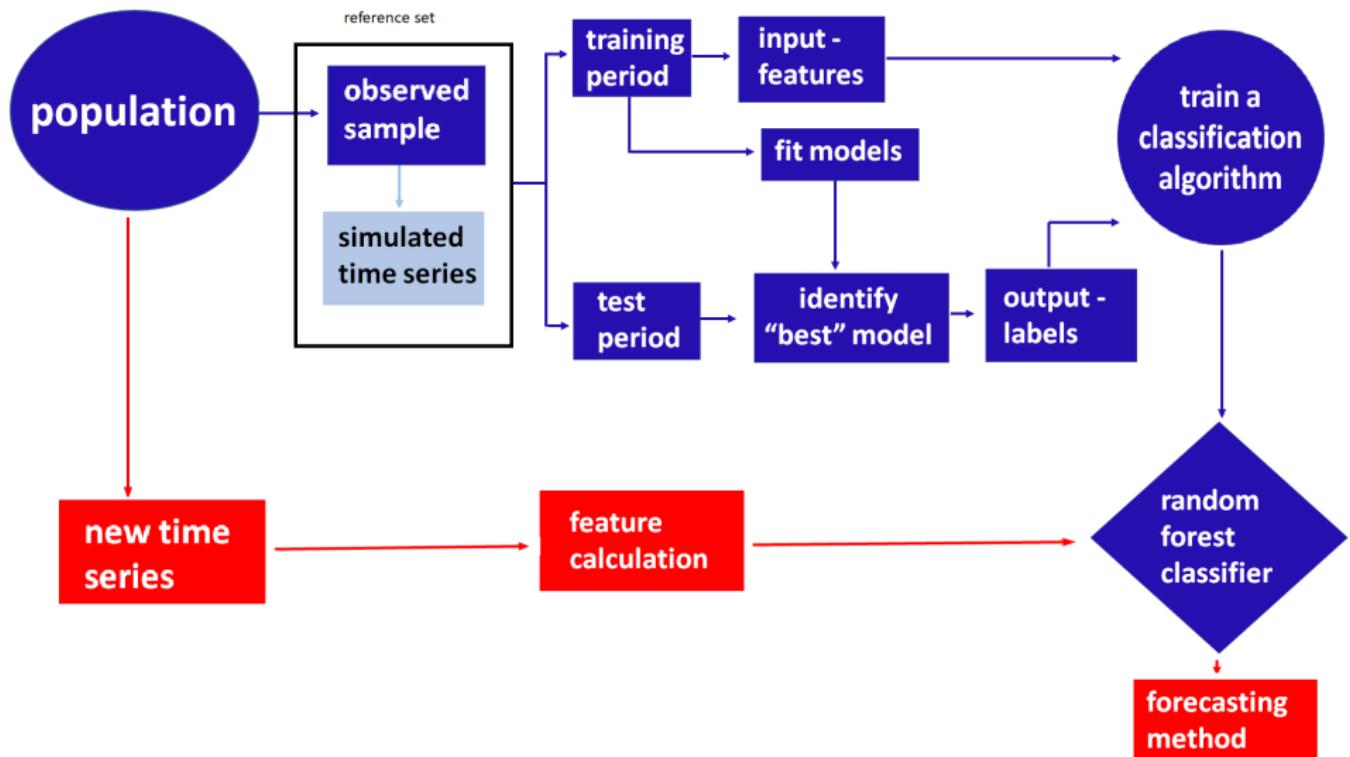
FFORMS: Feature-based FORecast Model Selection



FFORMS: Feature-based FORcast Model Selection



FFORMS: Feature-based FORecast Model Selection



Application to M competition data

Experiment 1

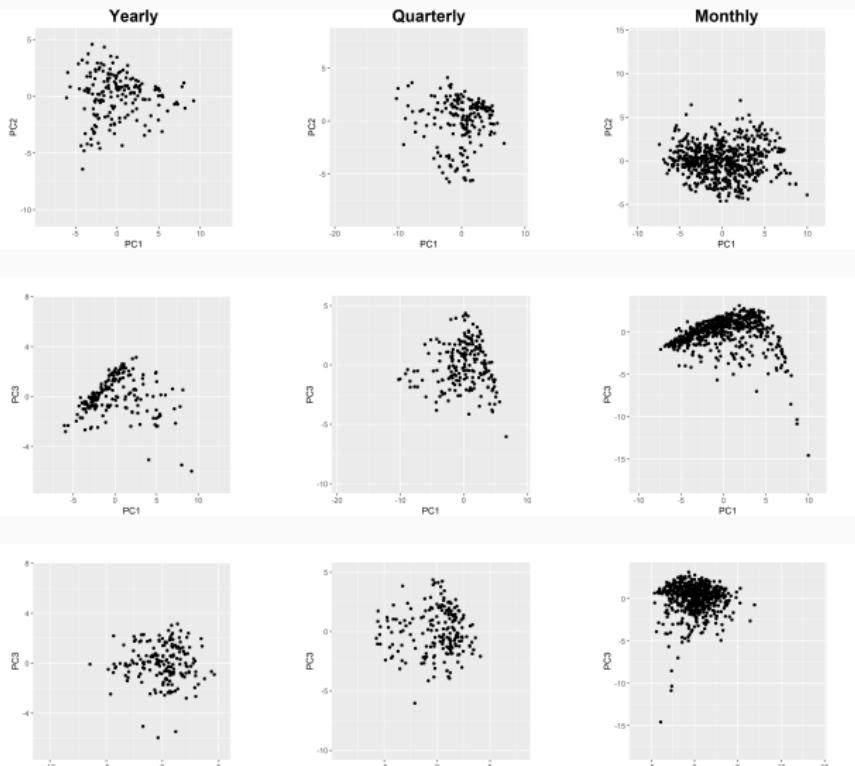
	Source	Y	Q	M
Observed series	M1	181	203	617
Simulated series		362000	406000	123400
New series	M3	645	756	1428

Experiment 2

	Source	Y	Q	M
Observed series	M3	645	756	1428
Simulated series		1290000	1512000	285600
New series	M1	181	203	617

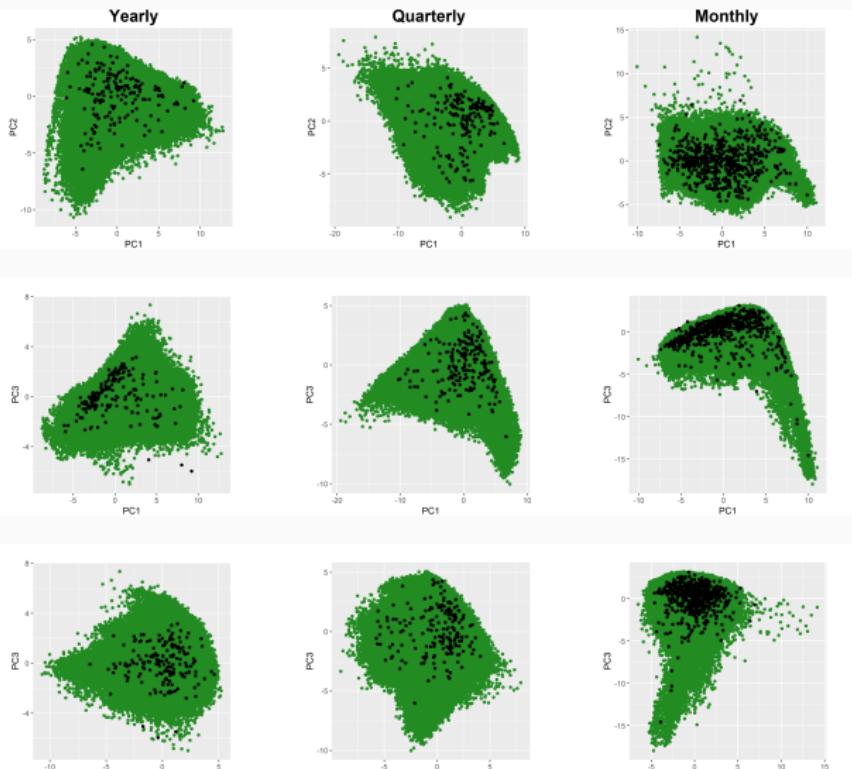
Experiment 1: Distribution of time series in PCA space

observed - M1



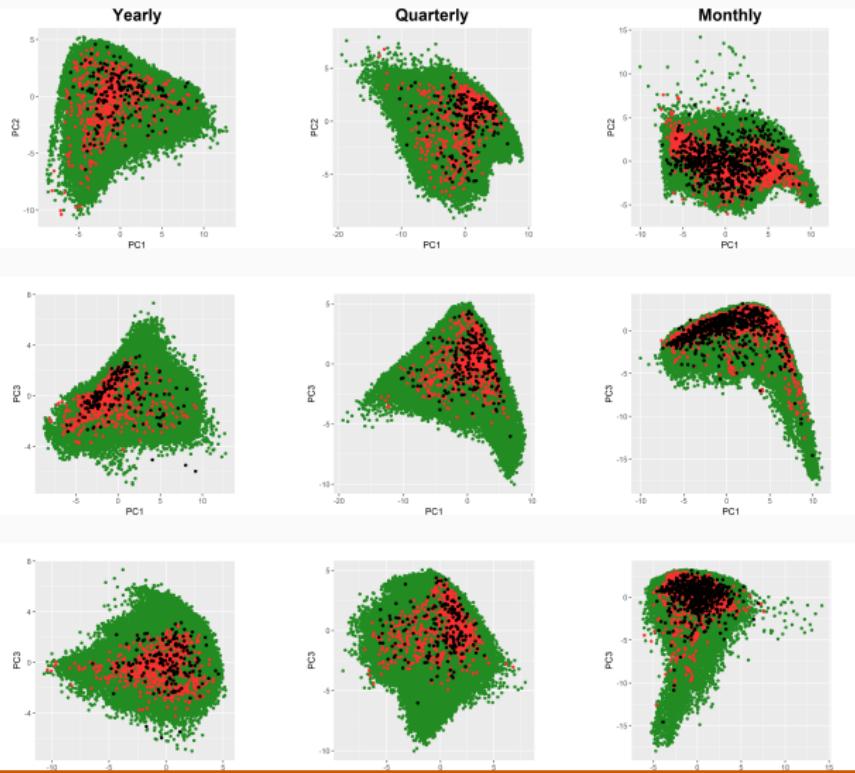
Experiment 1: Distribution of time series in PCA space

observed - M1 simulated



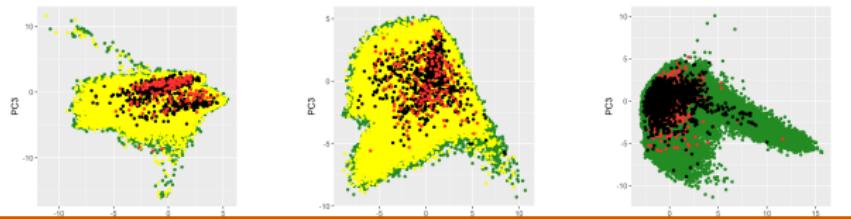
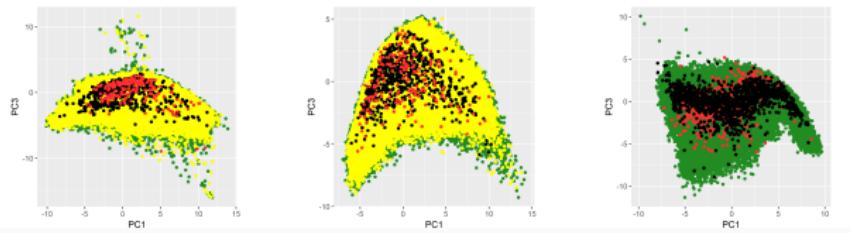
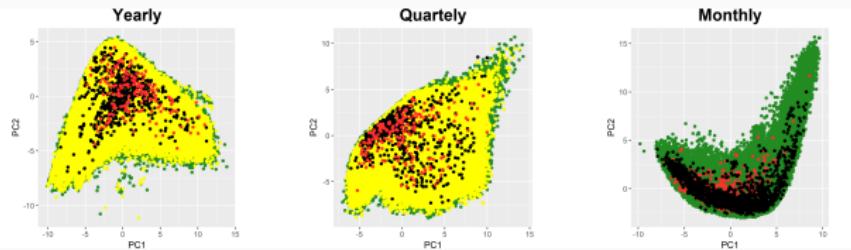
Experiment 1: Distribution of time series in PCA space

observed - M1 simulated new - M3

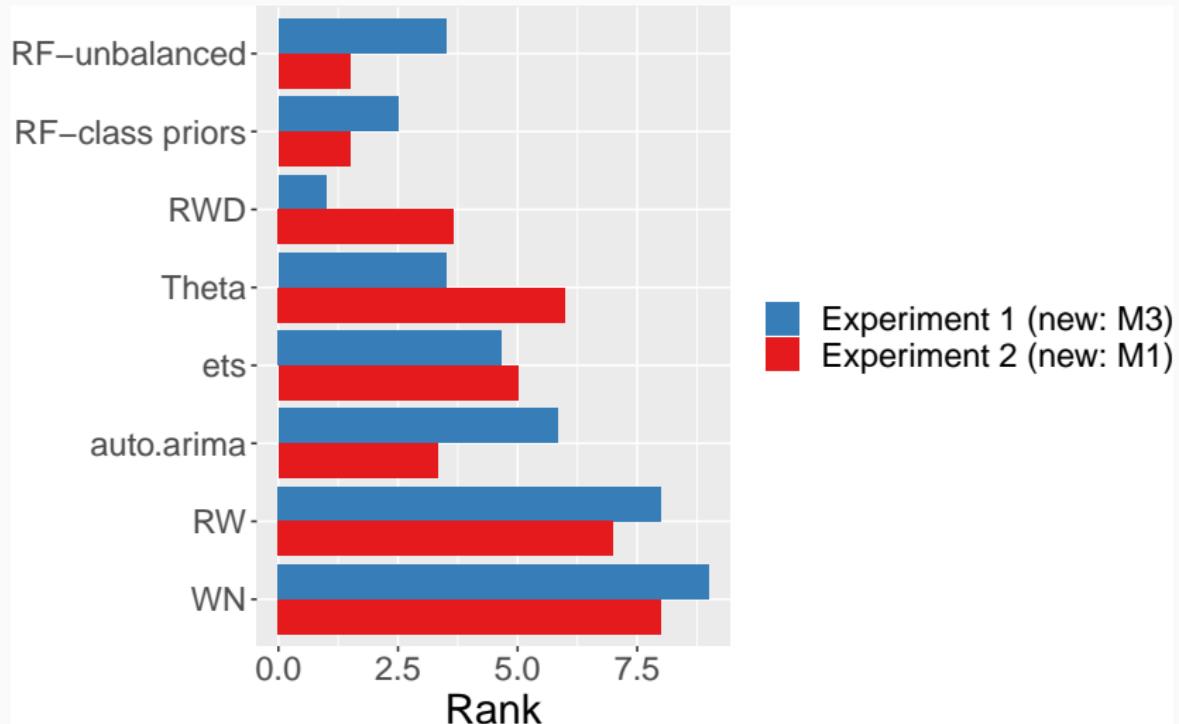


Experiment 2: Distribution of time series in PCA space

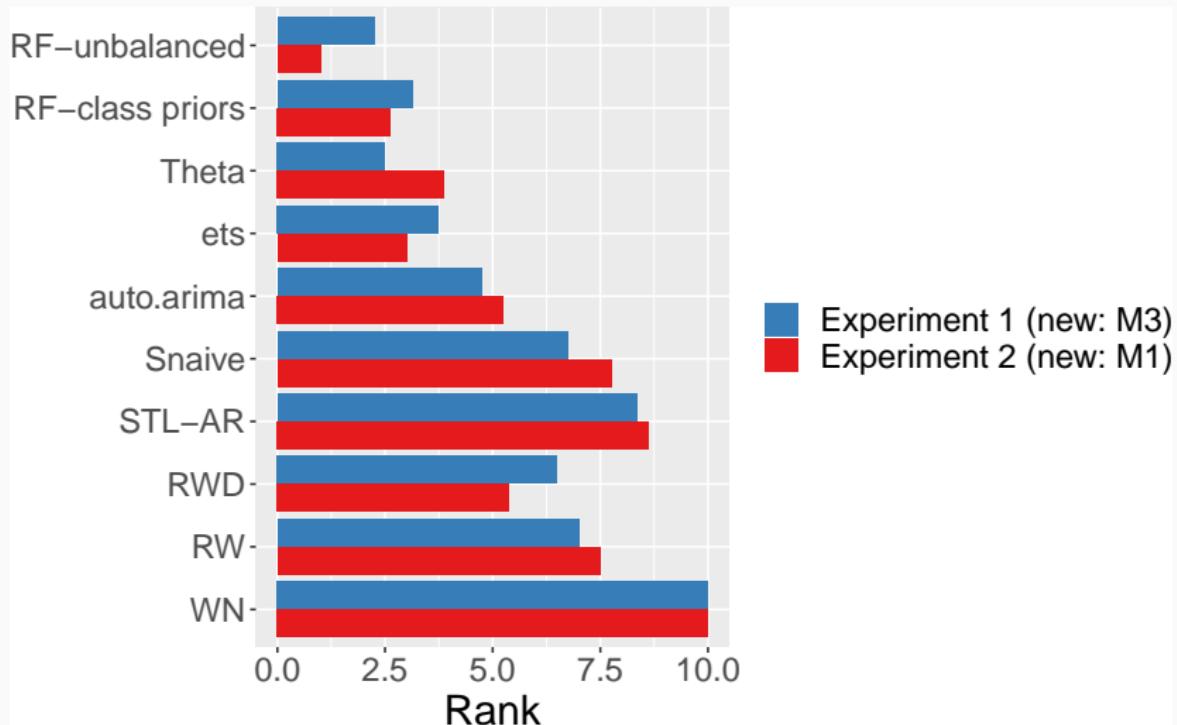
observed - M3 simulated subset new - M1



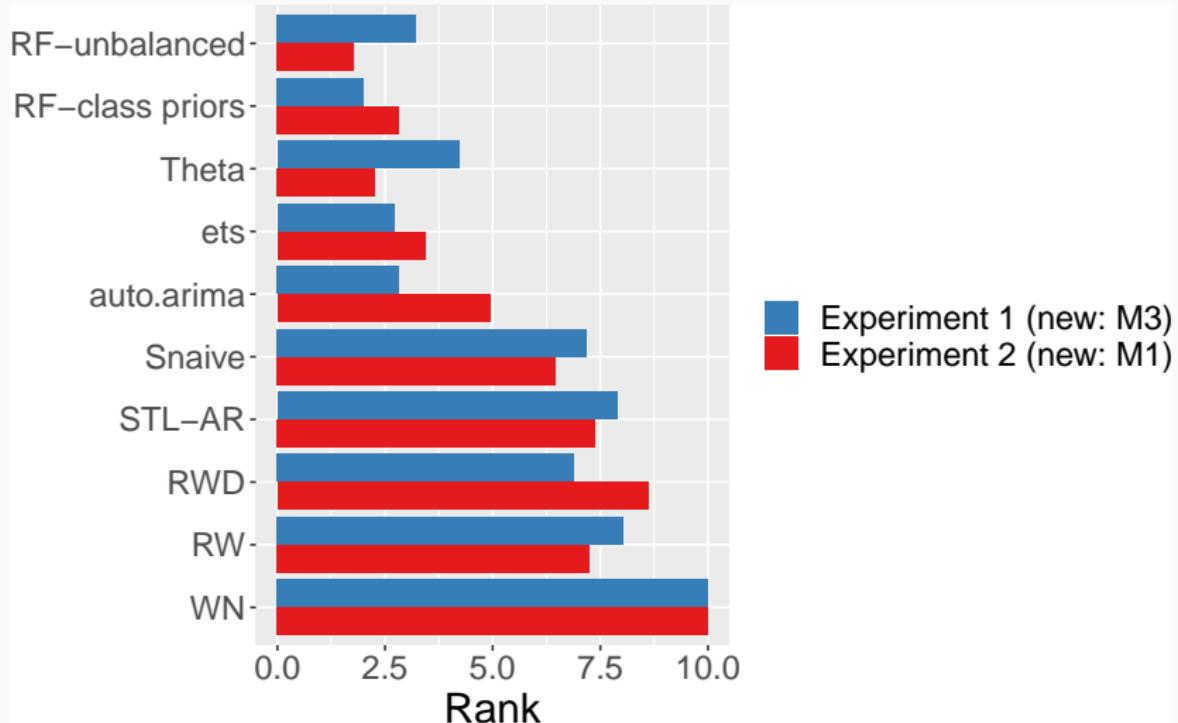
Results: Yearly



Results: Quarterly



Results: Monthly



FFORMA: Feature-based FOrecast Model Averaging

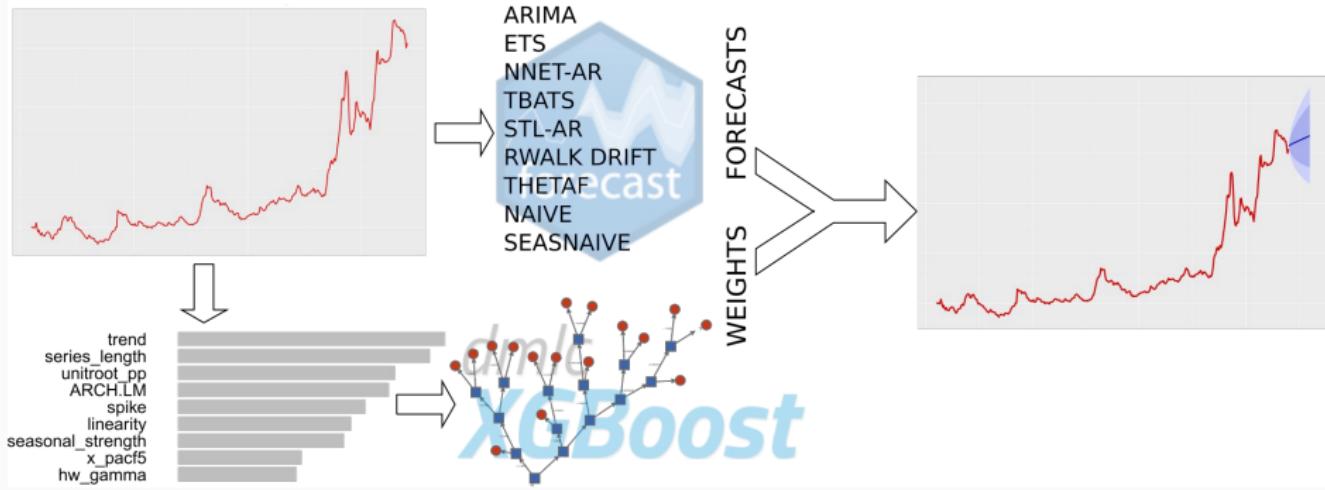
- Like FFORMS but using gradient boosted trees (xgboost) rather than random forest.
- Trained on temporal holdout version of M4 dataset, where size of test sets equal to required forecast horizons
- Optimization criterion: forecast accuracy not classification accuracy.
- Probability of each model being best is used to construct model weights for combination forecast.
- 5 days computing time.

FFORMA: Feature-based FOrecast Model Averaging

Models included

- 1 Naive
- 2 Seasonal naive
- 3 Random walk with drift
- 4 Theta method
- 5 ARIMA
- 6 ETS
- 7 TBATS
- 8 STL decomposition with AR for seasonally
adjusted series
- 9 Neural network autoregression

FFORMA: Feature-based FORcast Model Averaging

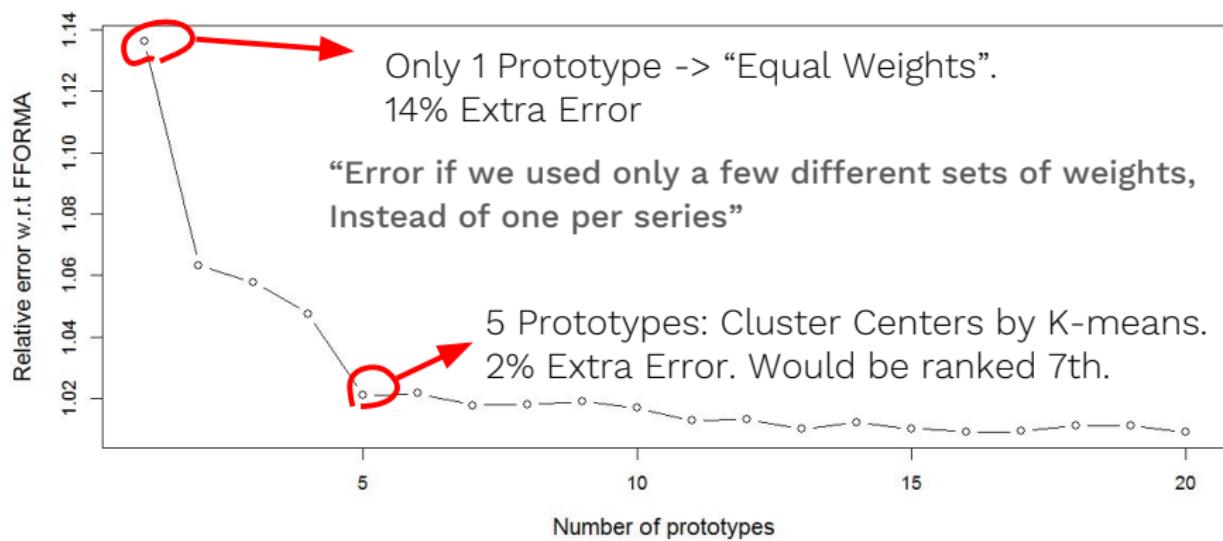


M4 competition results (based on average OWA)

1st	0.821
2nd	0.838 (FFORMA)
3rd	0.841

FFORMA: Feature-based FOrecast Model Averaging

Looking for Prototypes in the weights



FFORMA: Feature-based FOrecast Model Averaging

“Roughly Equal Weights”. 40000 Series in M4

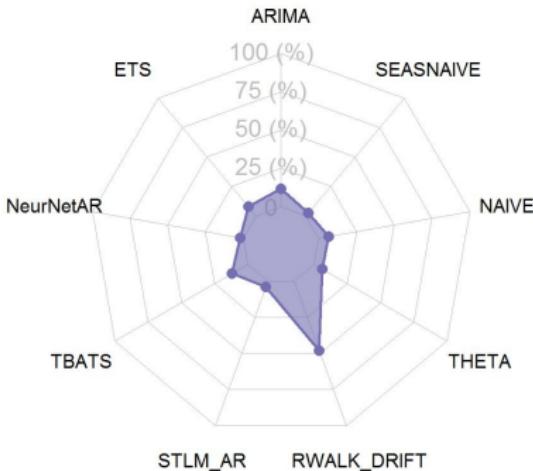
Weights of Prototype I



FFORMA: Feature-based FOrecast Model Averaging

“Mostly RandomWalk Drift”. 20000 Series in M4

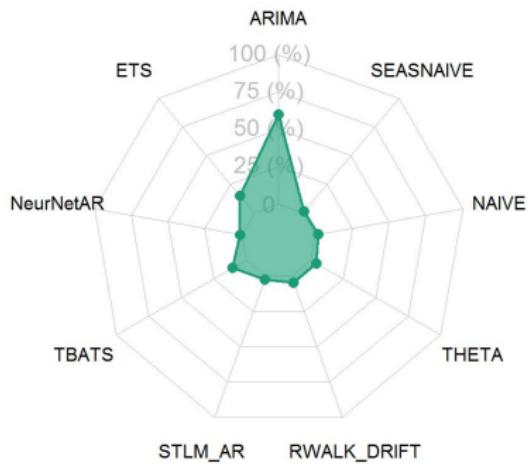
Weights of Prototype II



FFORMA: Feature-based FOrecast Model Averaging

“Mostly ARIMA”. 16000 Series in M4

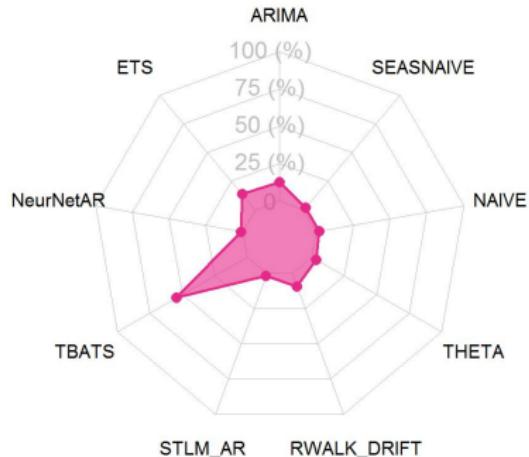
Weights of Prototype III



FFORMA: Feature-based FOrecast Model Averaging

“Mostly TBATS”. 13000 Series in M4

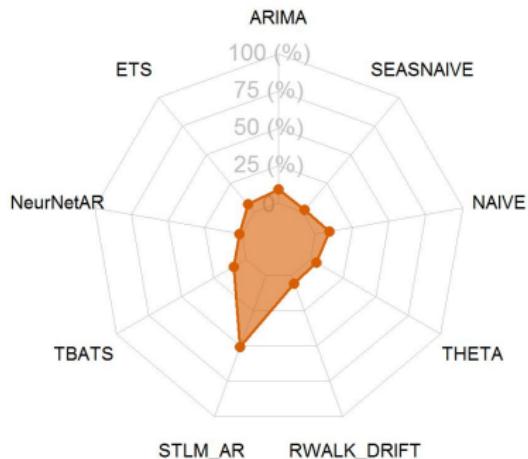
Weights of Prototype IV



FFORMA: Feature-based FOrecast Model Averaging

“Mostly STLM-AR”. 8000 Series in M4

Weights of Prototype V



Acknowledgments



Dilini Talagala



Earo Wang



Kate Smith-Miles



George Athanasopoulos



Thiyanga Talagala



Yanfei Kang



Mitchell O'Hara-Wild



Pablo Montero-Manso

Papers

Available from robjhyndman.com