

# Feature-based time series analysis

Rob J Hyndman

27 September 2019

# Outline

1 Visualization

2 R packages

3 Anomaly detection

4 Forecast model selection

5 Forecast model averaging

# Outline

1 Visualization

2 R packages

3 Anomaly detection

4 Forecast model selection

5 Forecast model averaging

# M3 competition



ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

[www.elsevier.com/locate/ijforecast](http://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 4

# M3 competition



ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

*international journal  
of forecasting*

[www.elsevier.com/locate/ijforecast](http://www.elsevier.com/locate/ijforecast)



petition: results, conclusions a

Spyros Makridakis, Michèle Hibon\*

EAD, Boulevard de Constance, 77305 Fontainebleau, Fr

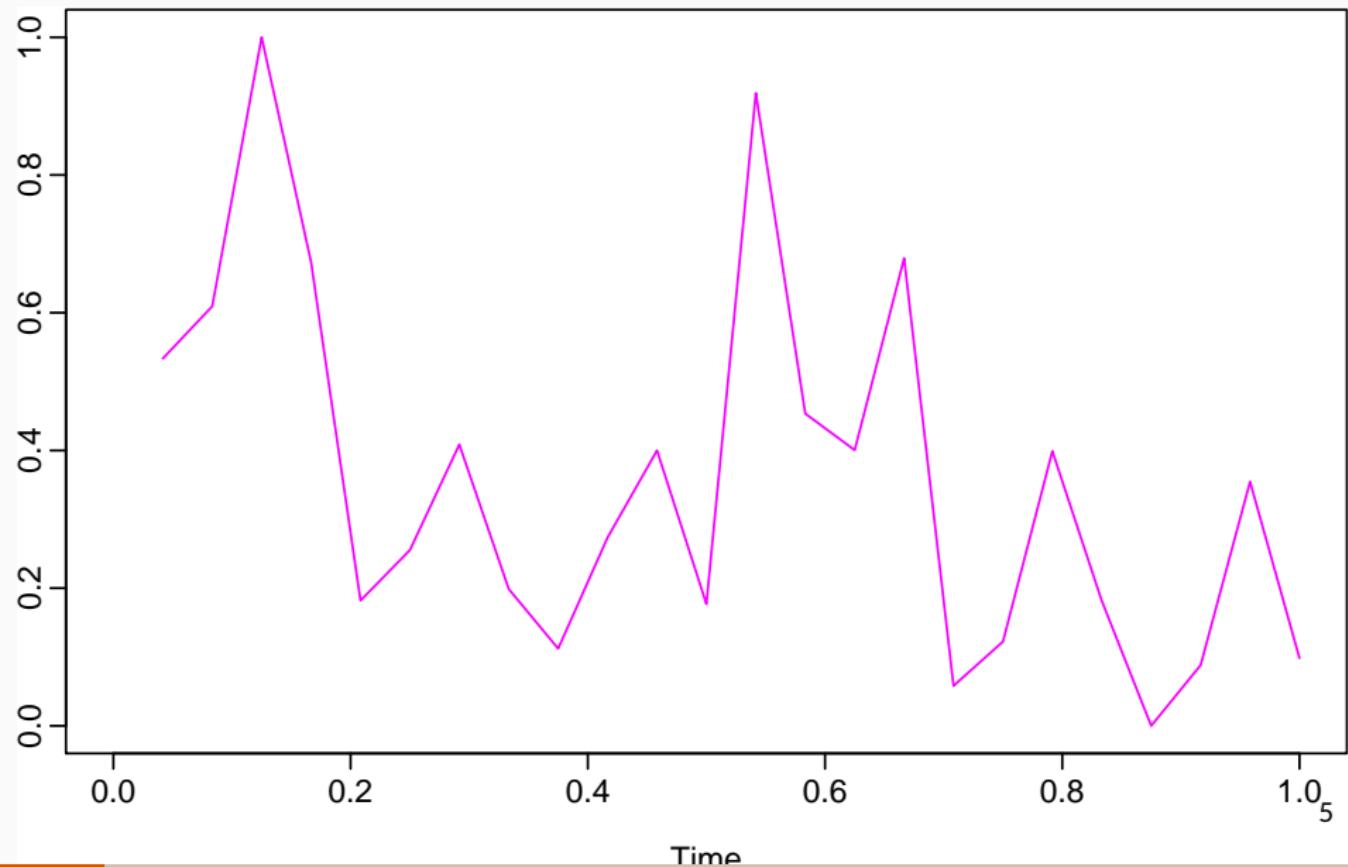
Abst



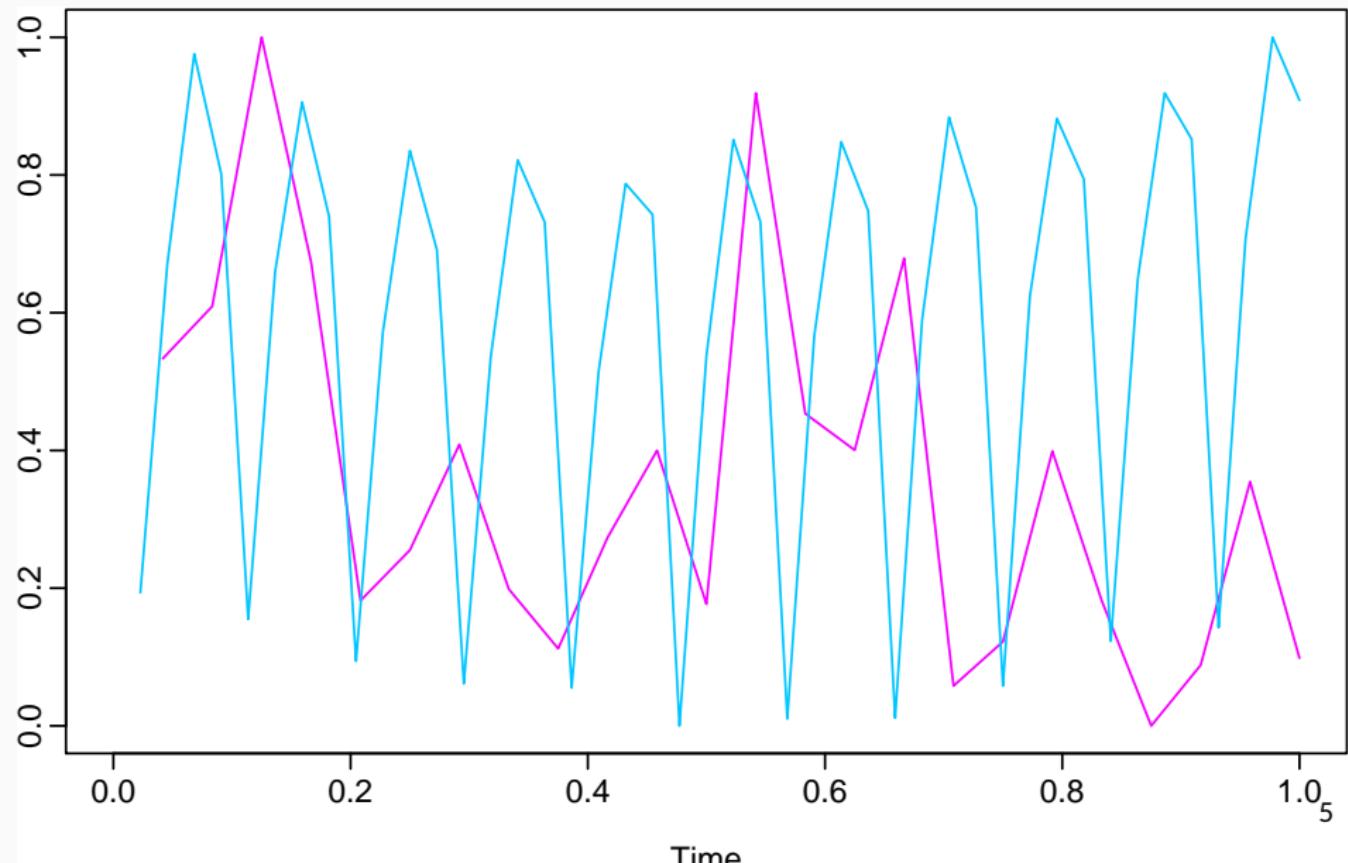
This paper describes the M3-Competition, the latest of the M-Competitions. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results/conclusions with those of the previous two M-Competitions as well as with those of 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

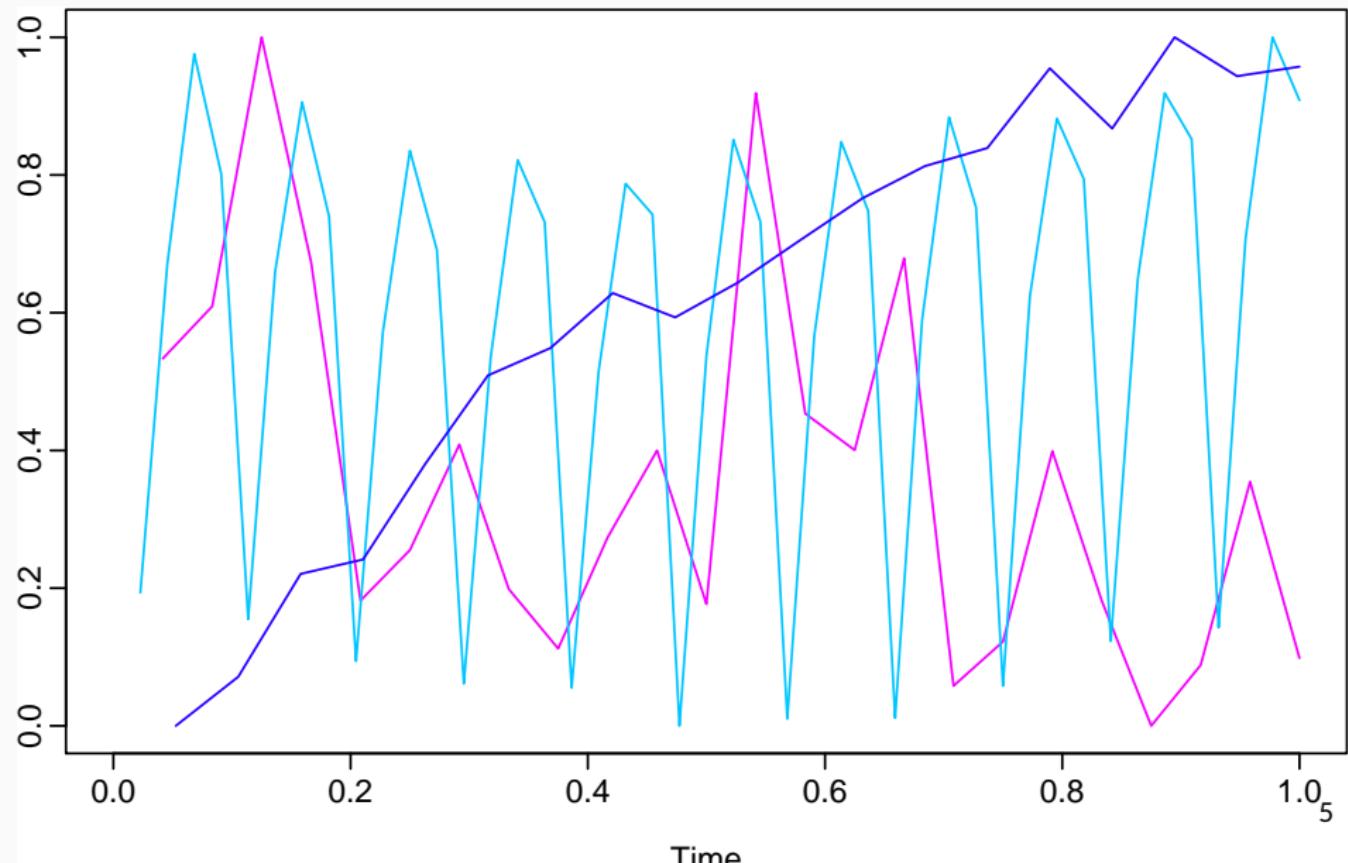
# How to plot lots of time series?



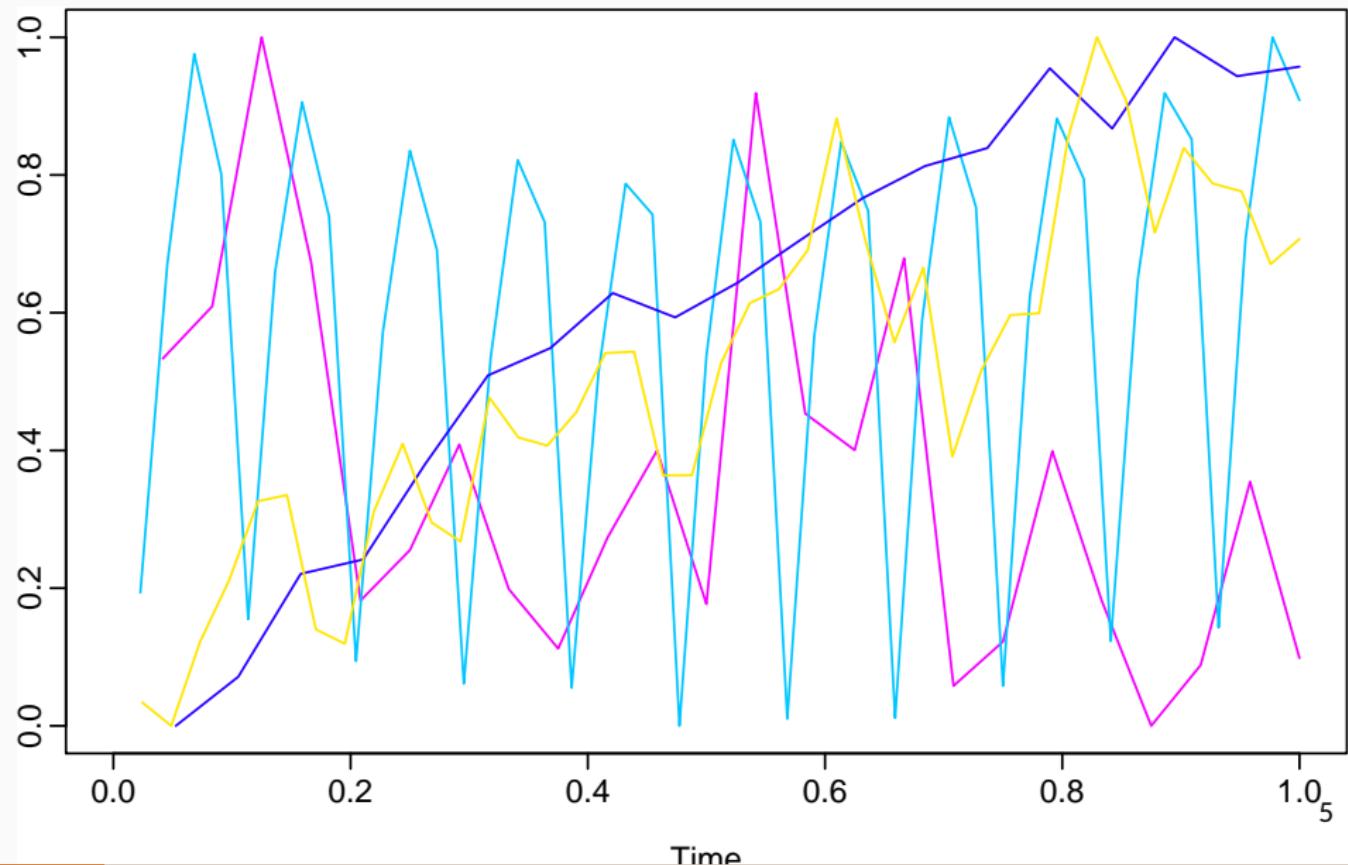
# How to plot lots of time series?



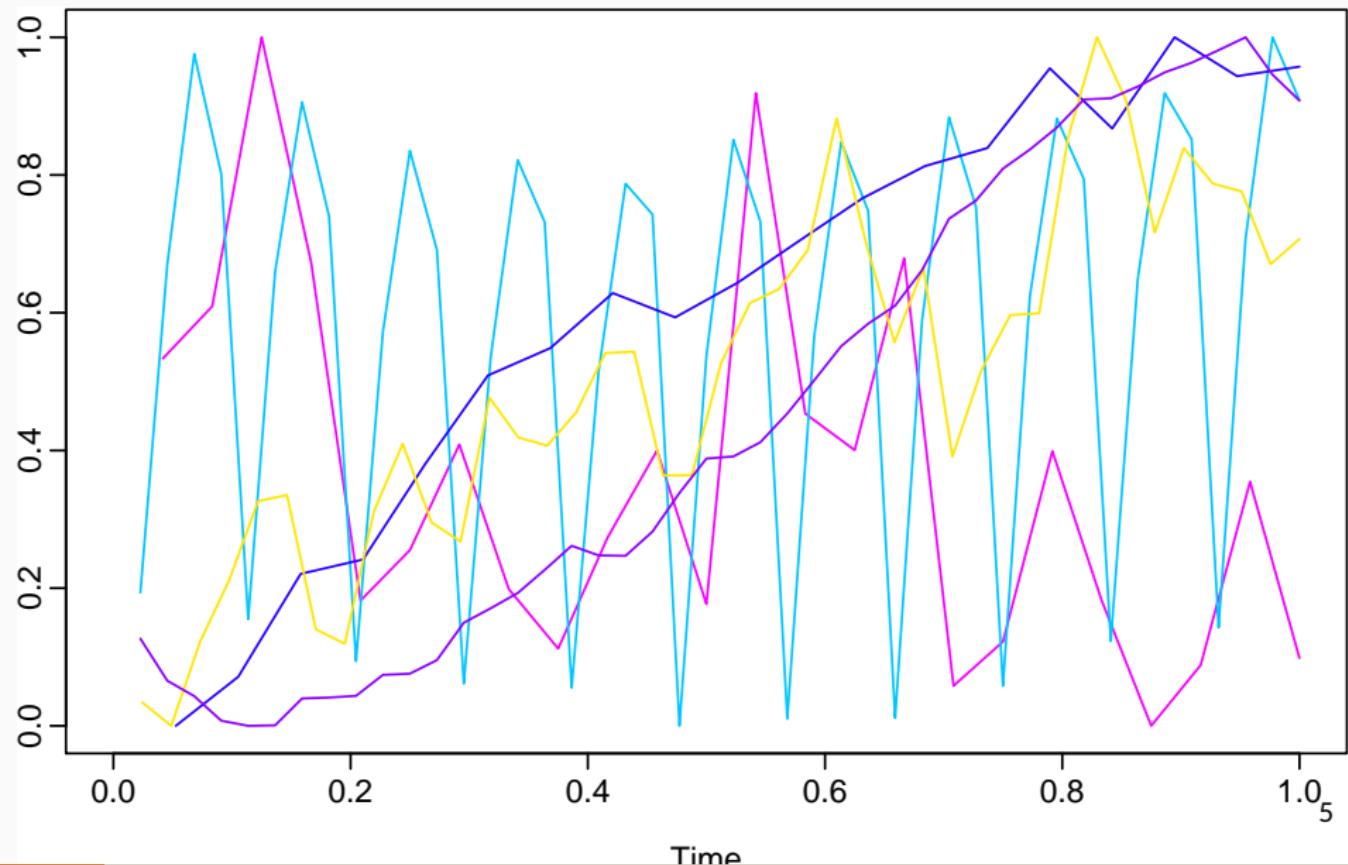
# How to plot lots of time series?



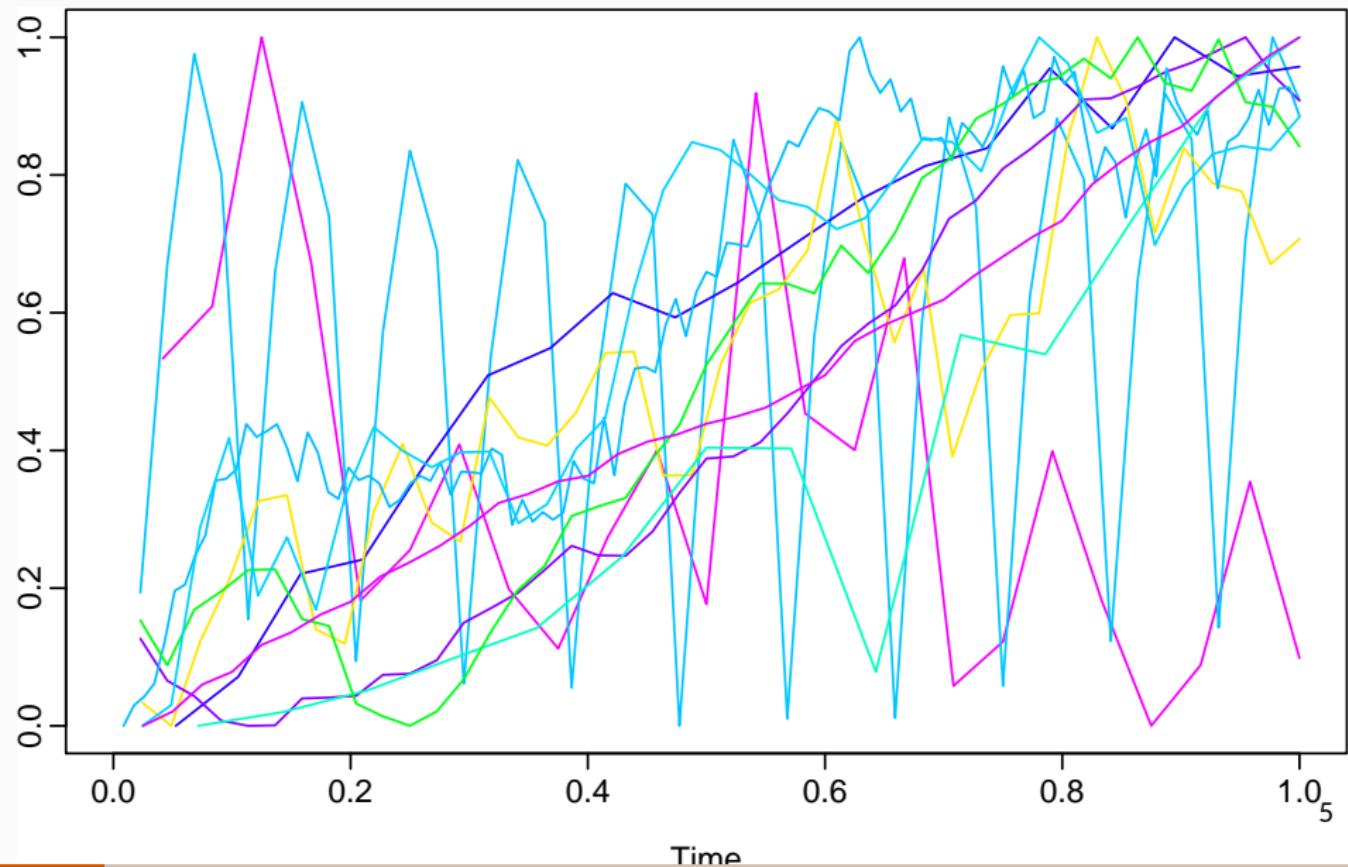
# How to plot lots of time series?



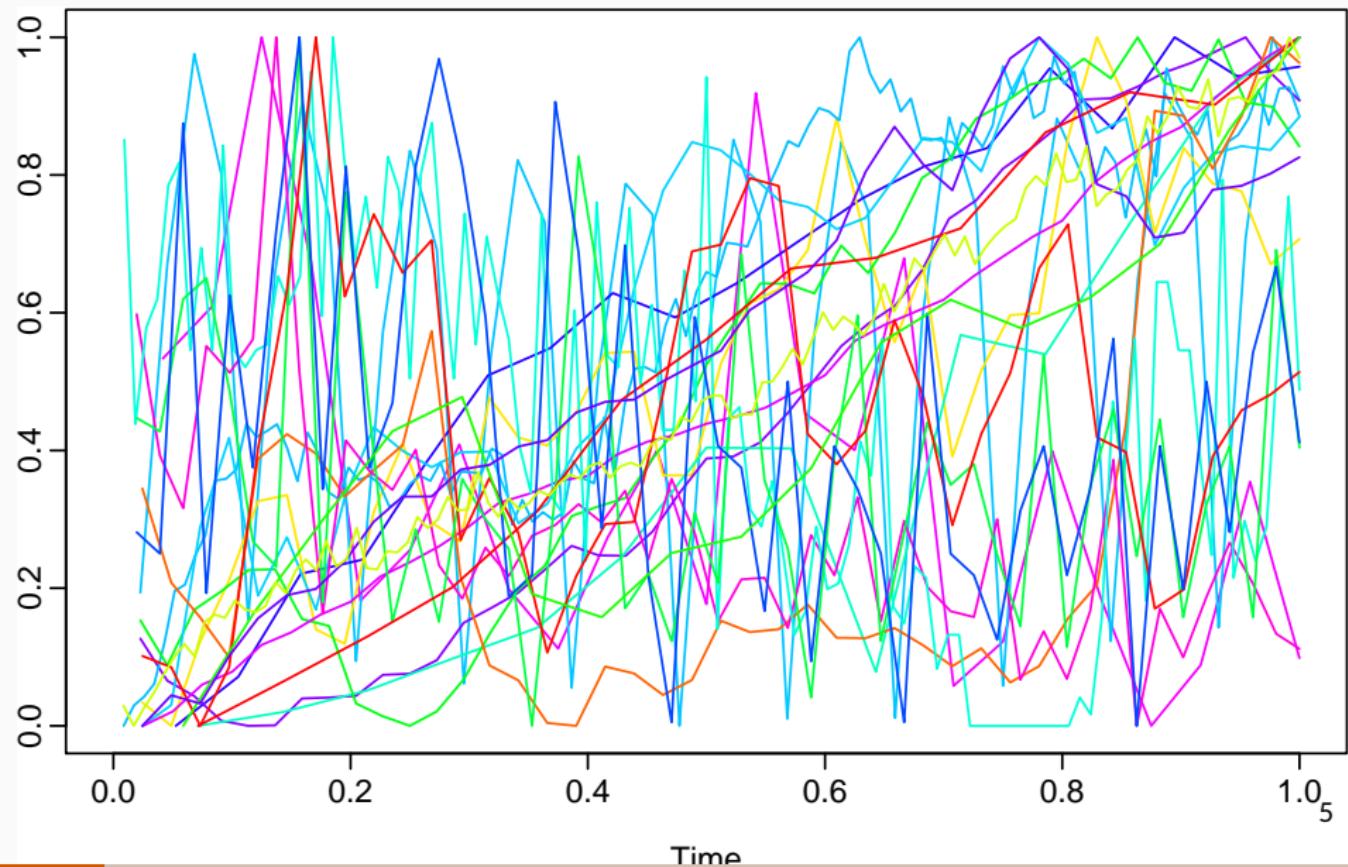
# How to plot lots of time series?



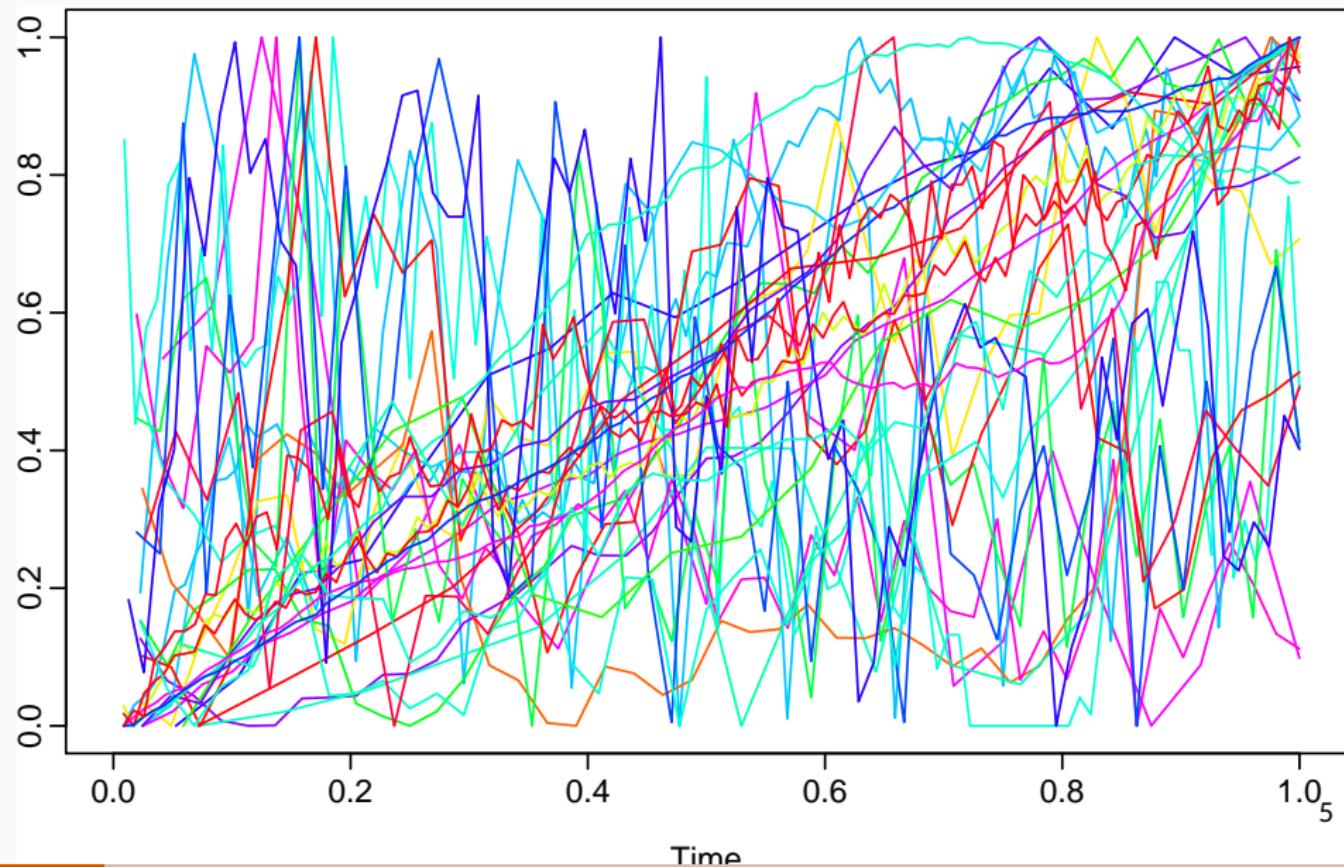
# How to plot lots of time series?



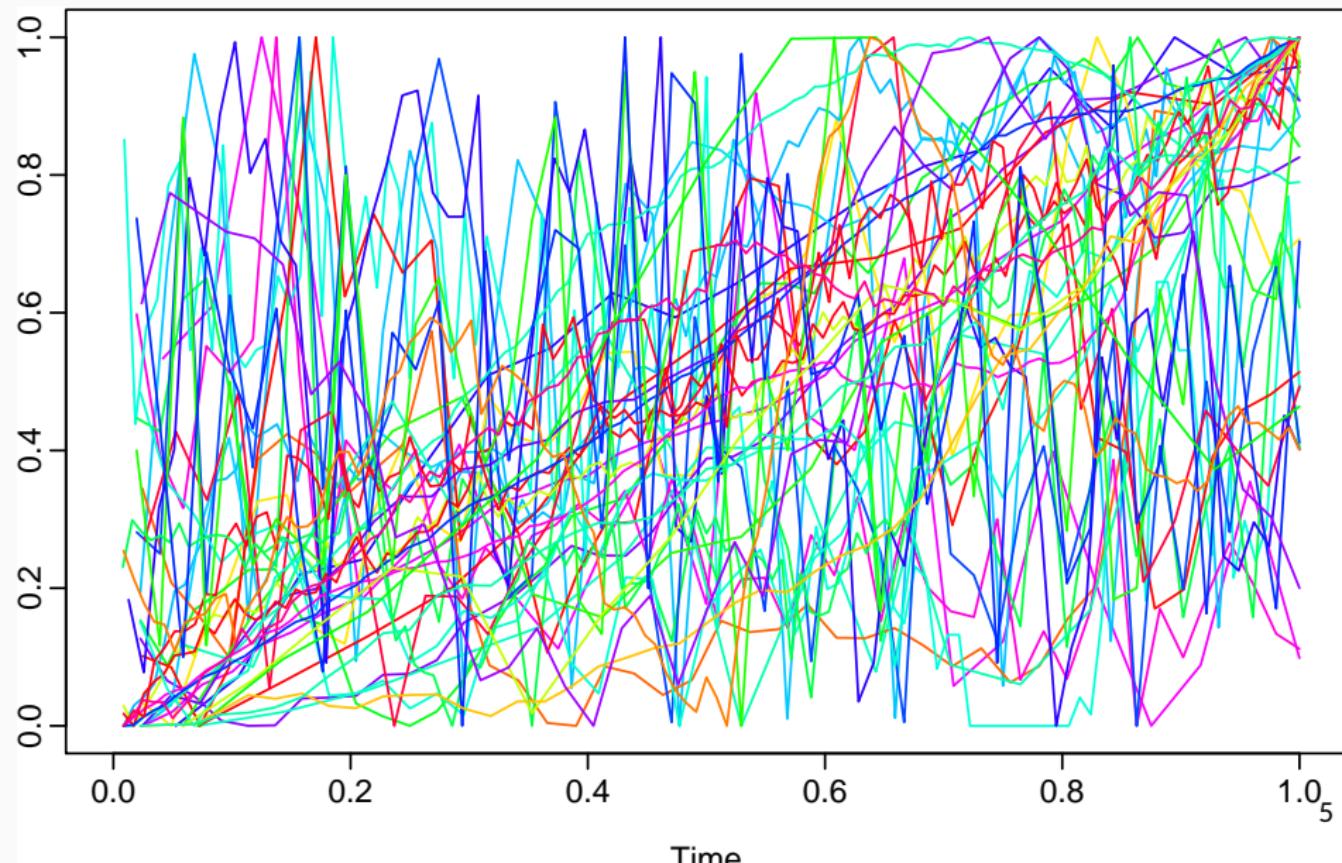
# How to plot lots of time series?



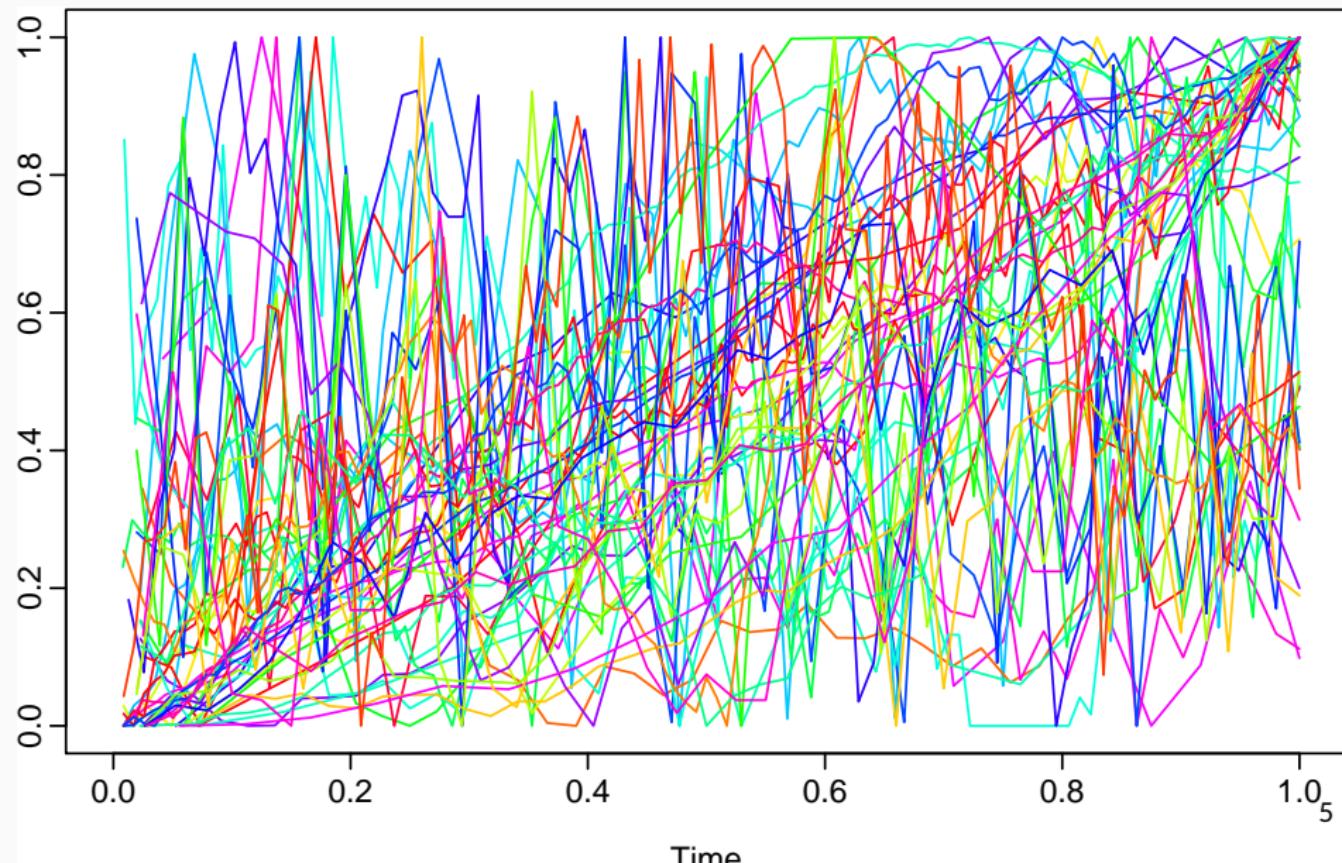
# How to plot lots of time series?



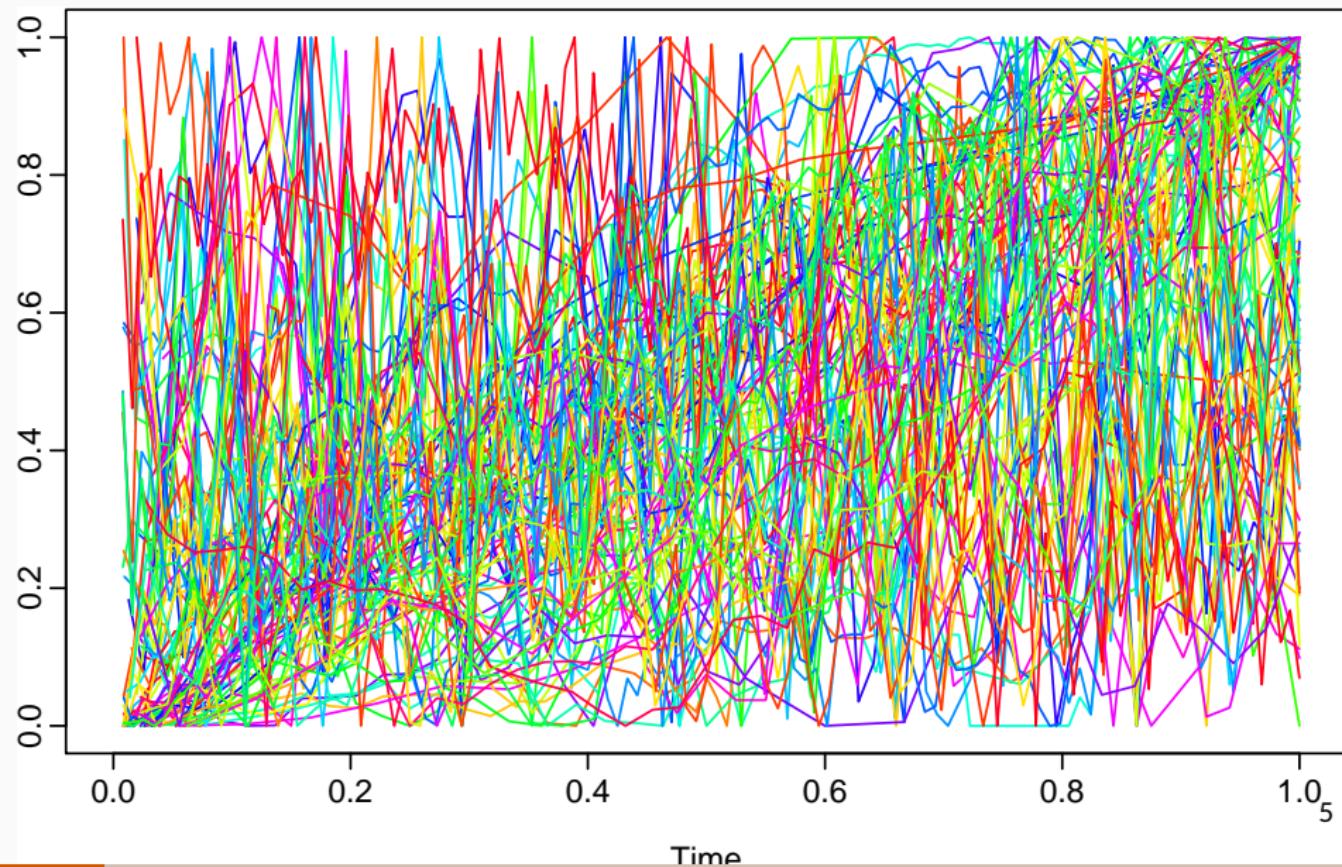
# How to plot lots of time series?



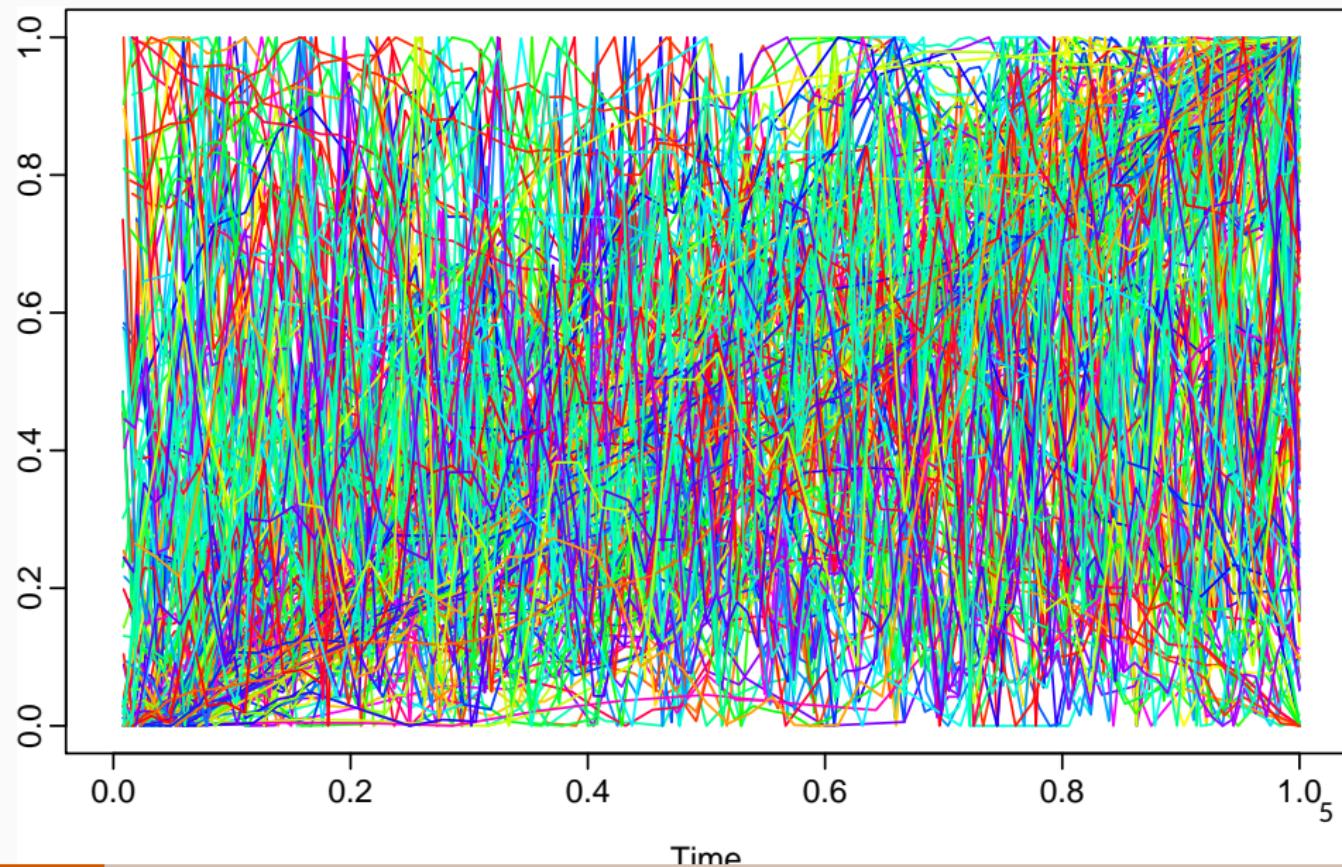
# How to plot lots of time series?



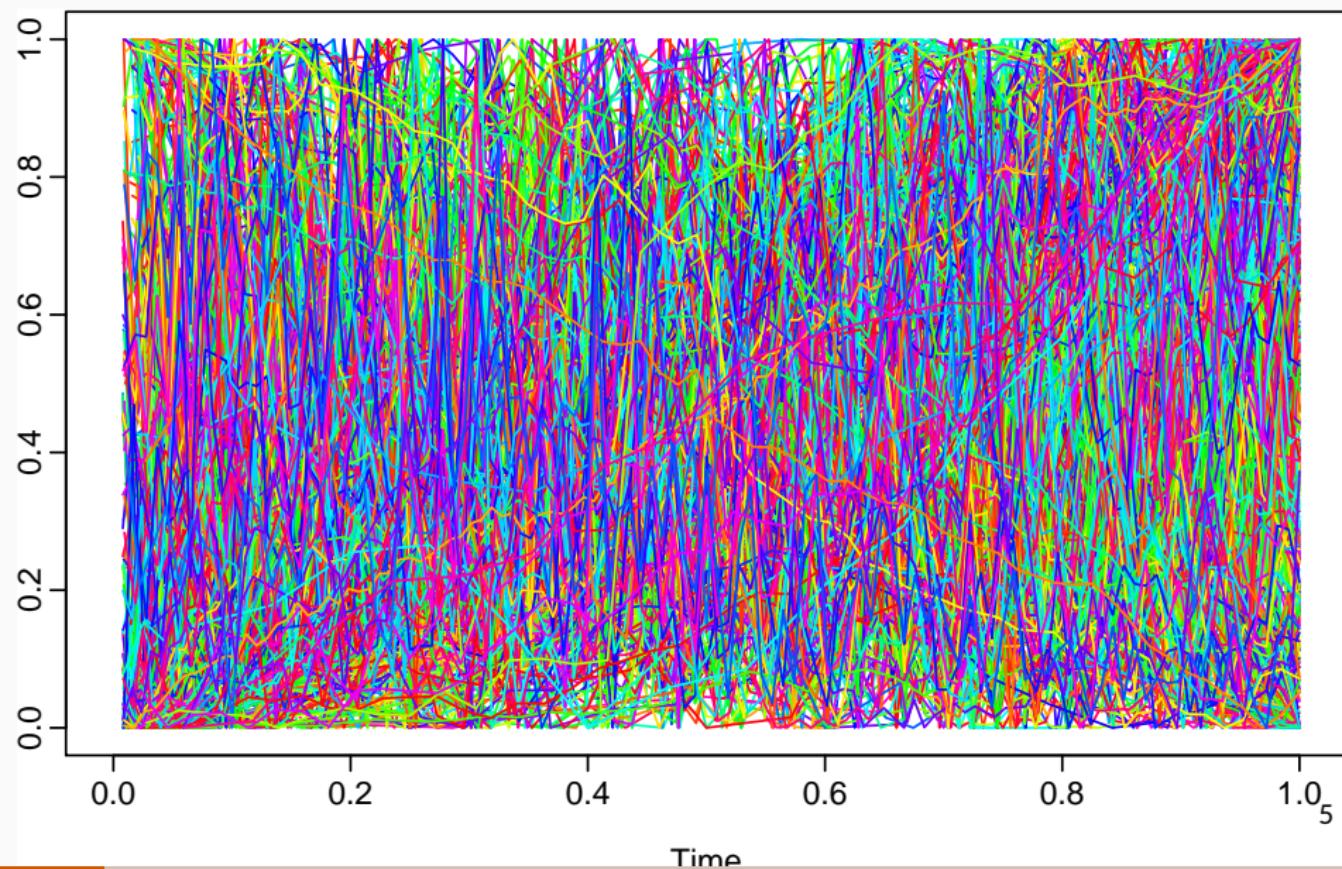
# How to plot lots of time series?



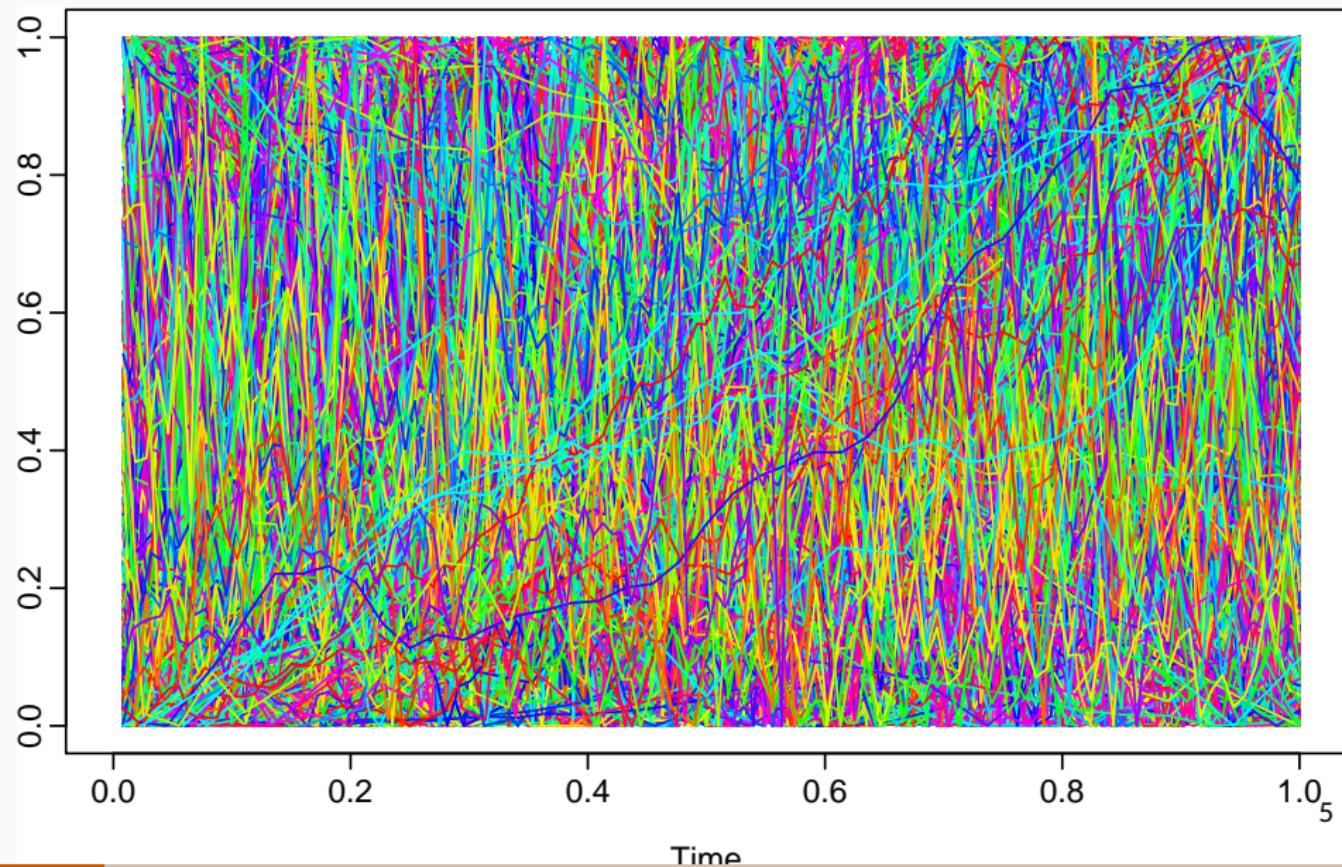
# How to plot lots of time series?



# How to plot lots of time series?



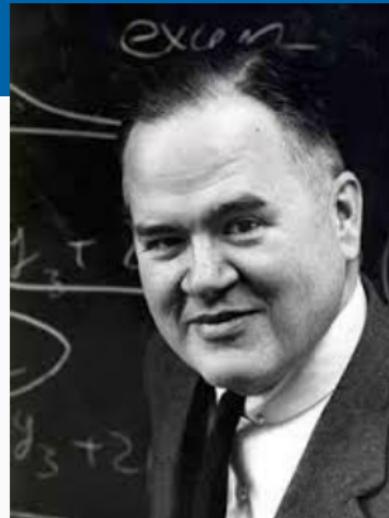
# How to plot lots of time series?



# Key idea

## Cognostics

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

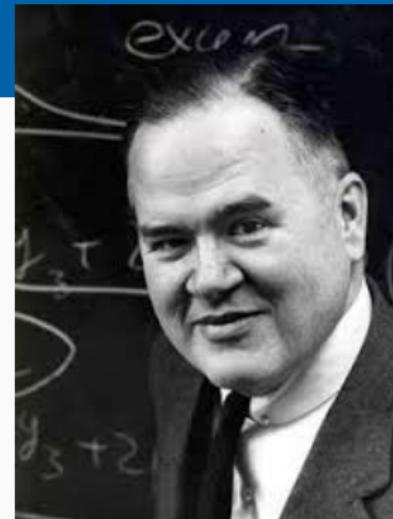


*John W Tukey*

# 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.

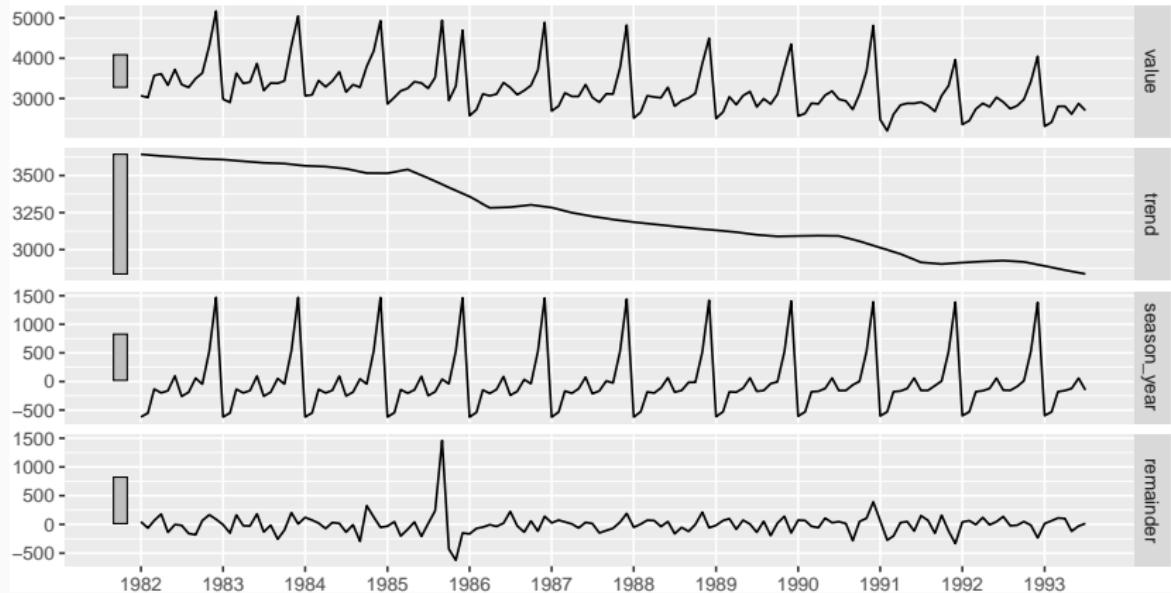
# An STL decomposition: N2096

$$Y_t = S_t + T_t + R_t$$

$S_t$  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

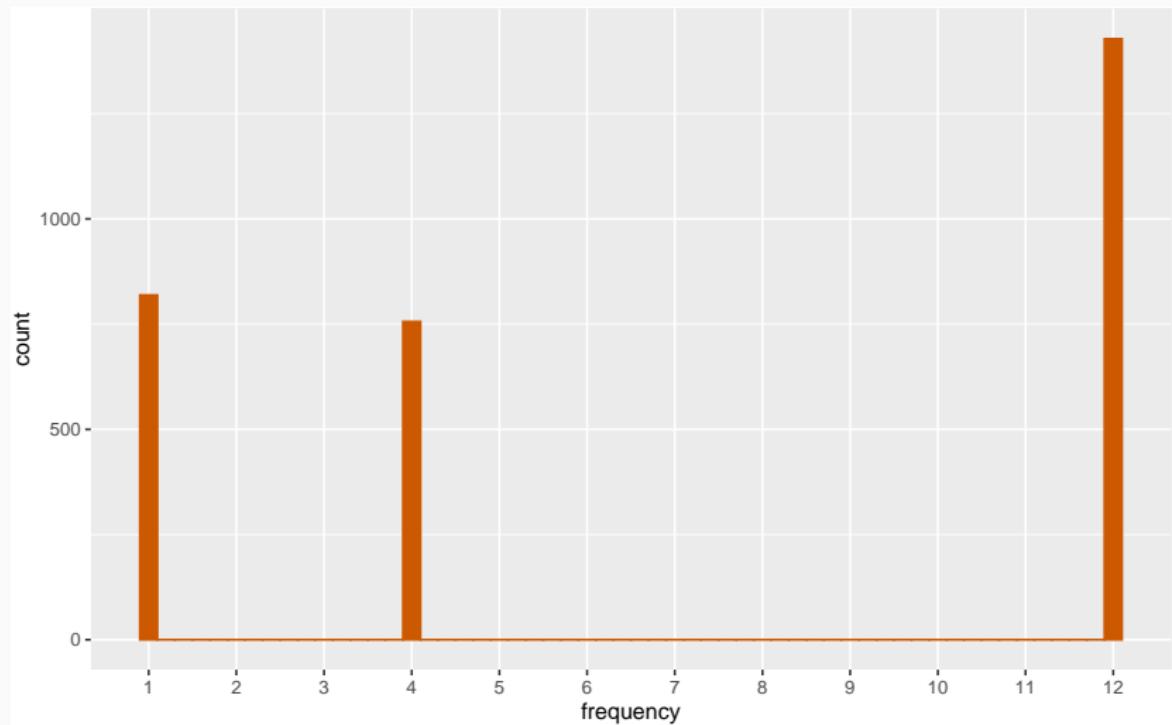
For series with different lengths, scales, domains, etc., we need features that are:

- scale-independent
- ergodic

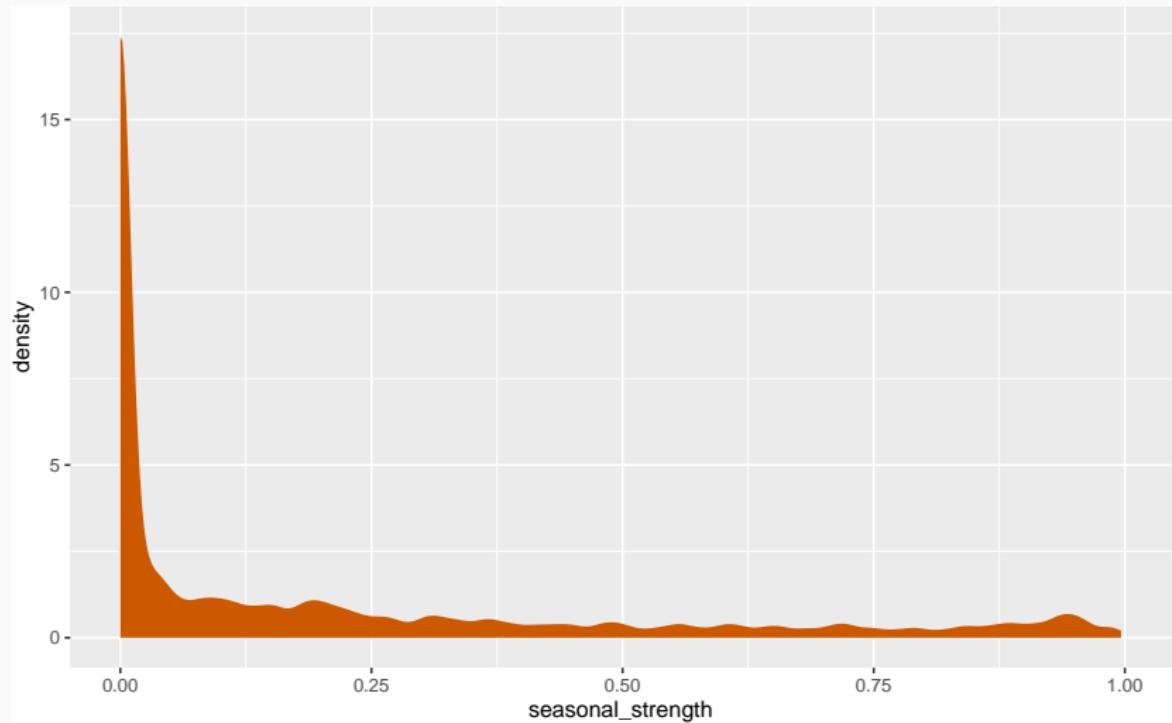
- 1 Seasonal period
- 2 Strength of seasonality
- 3 Strength of trend
- 4 First autocorrelation of STL remainder series
- 5 Spectral entropy
- 6 Optimal MLE Box-Cox transformation of data

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

# Distribution of Period for M3

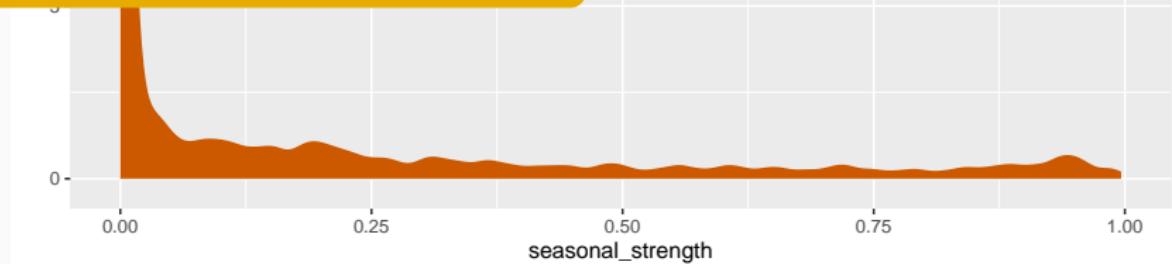
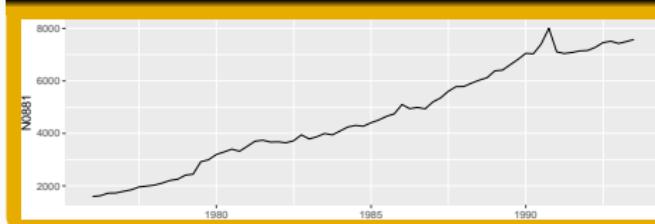


# Distribution of Seasonality for M3

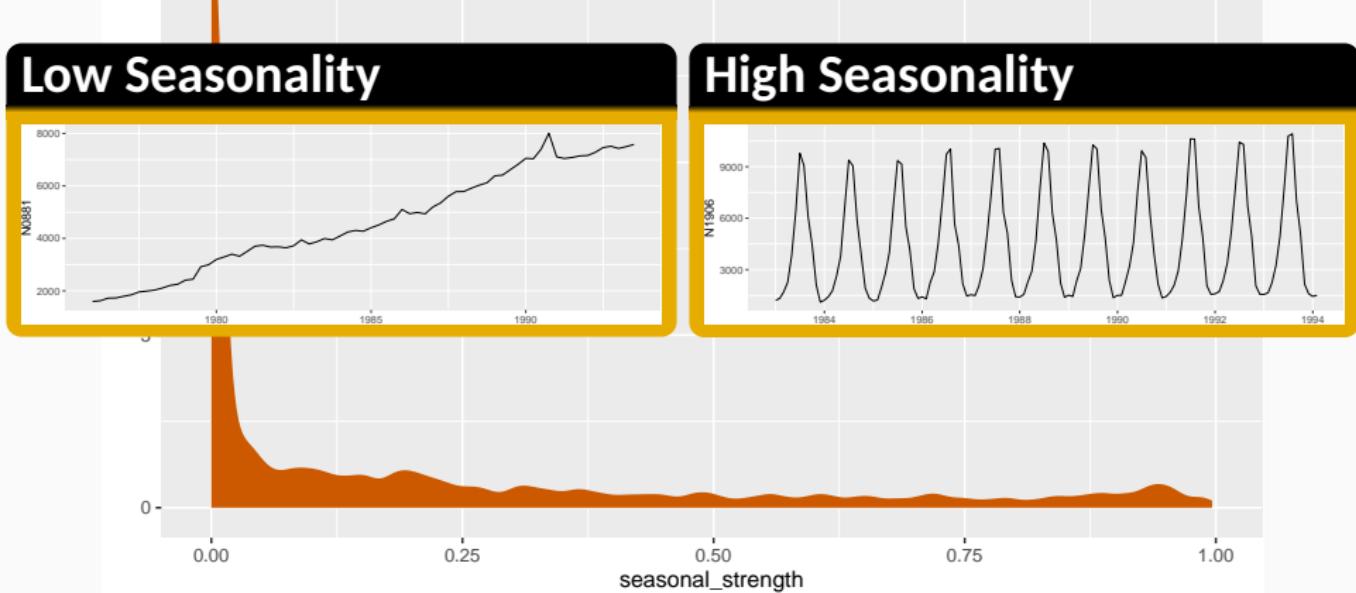


# Distribution of Seasonality for M3

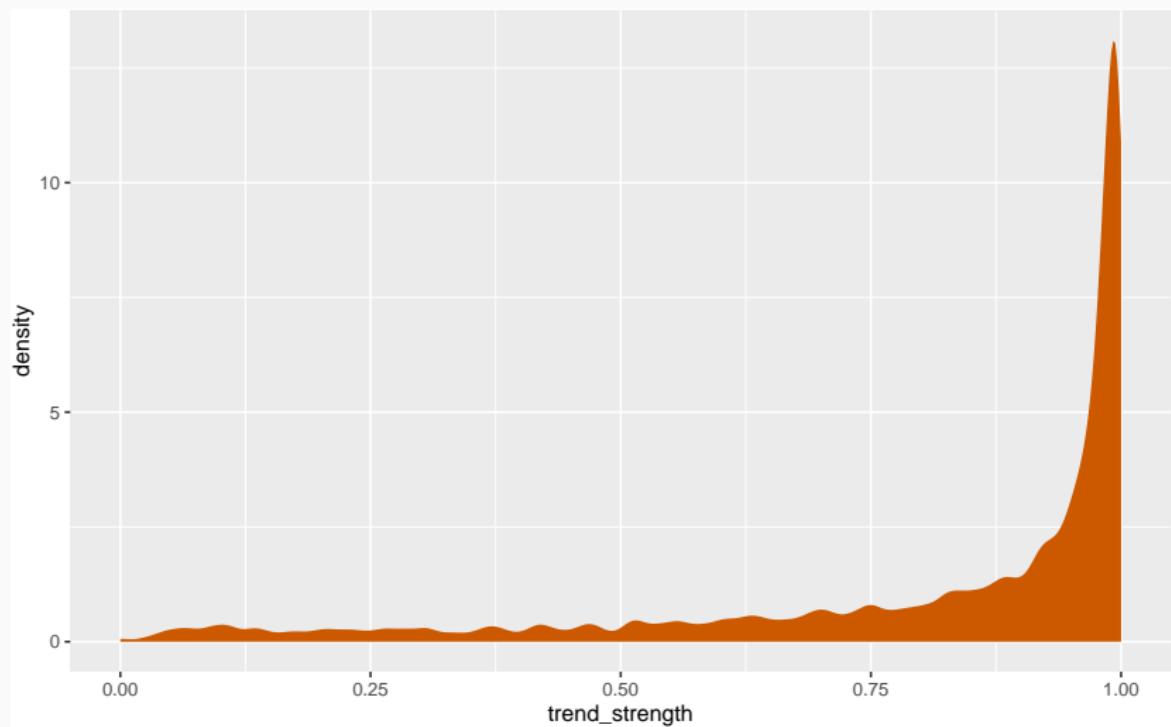
Low Seasonality



# Distribution of Seasonality for M3

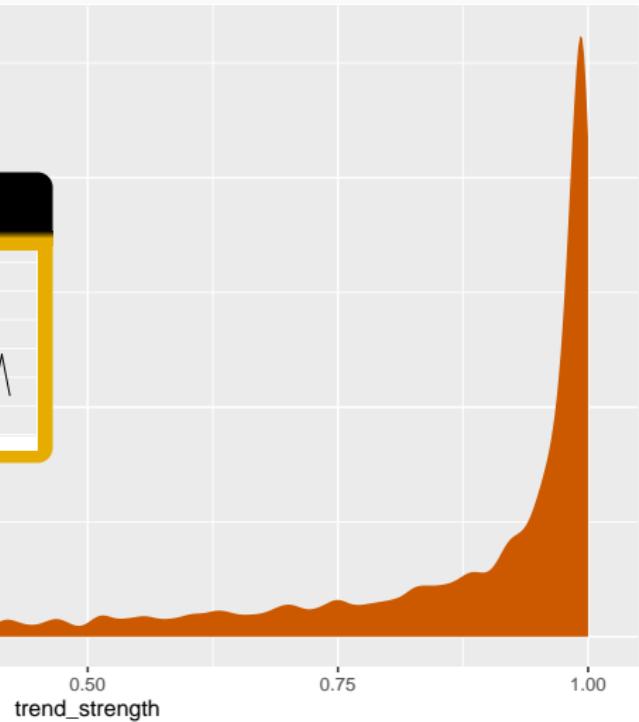
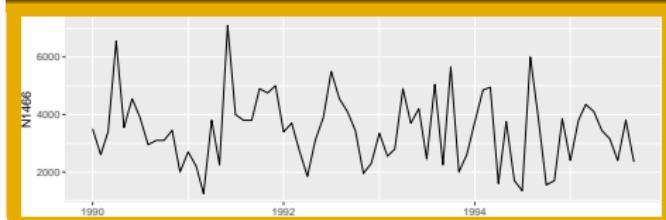


# Distribution of Trend for M3

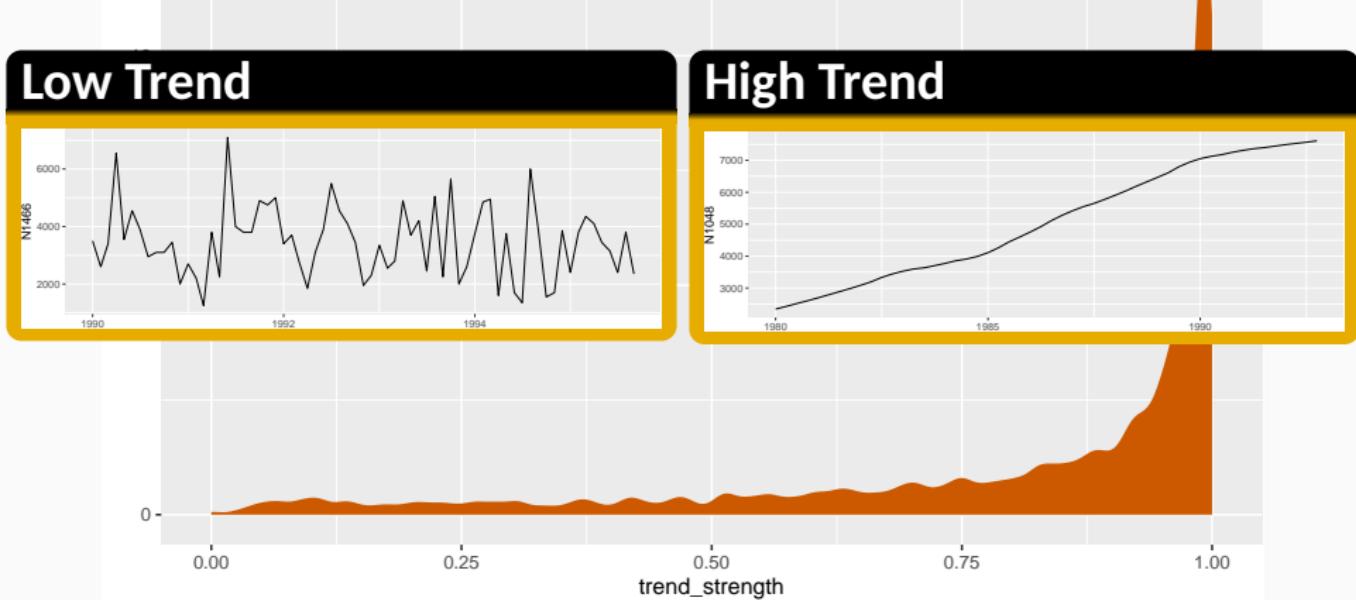


# Distribution of Trend for M3

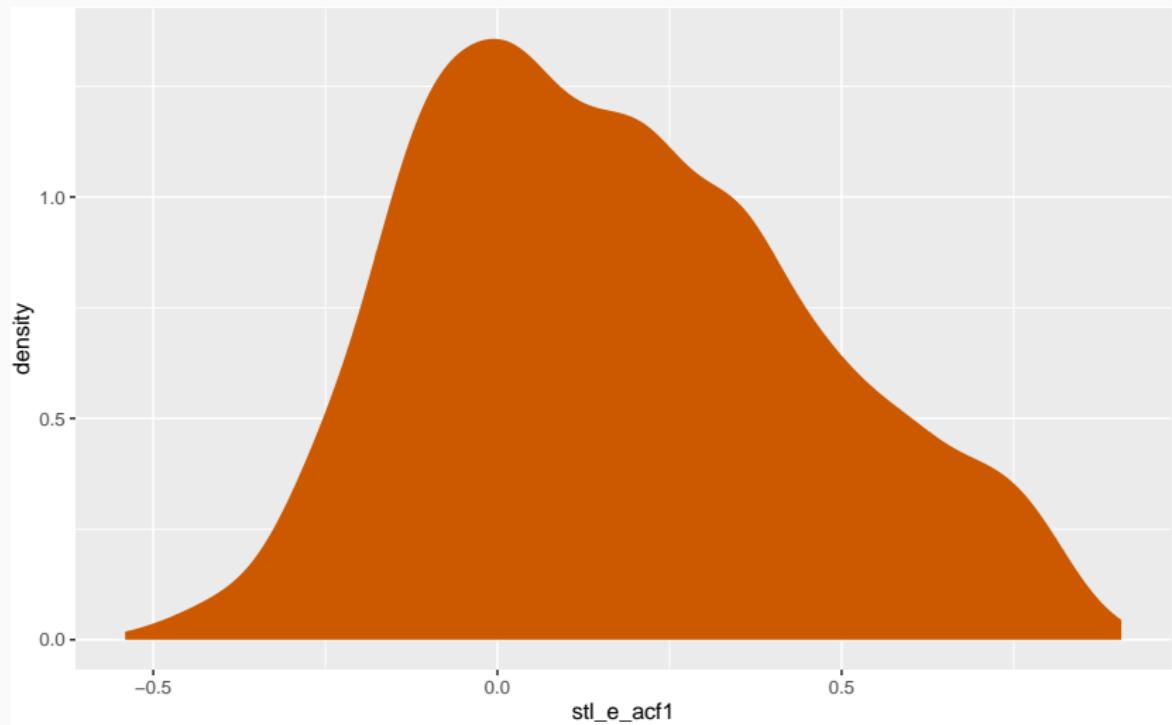
Low Trend



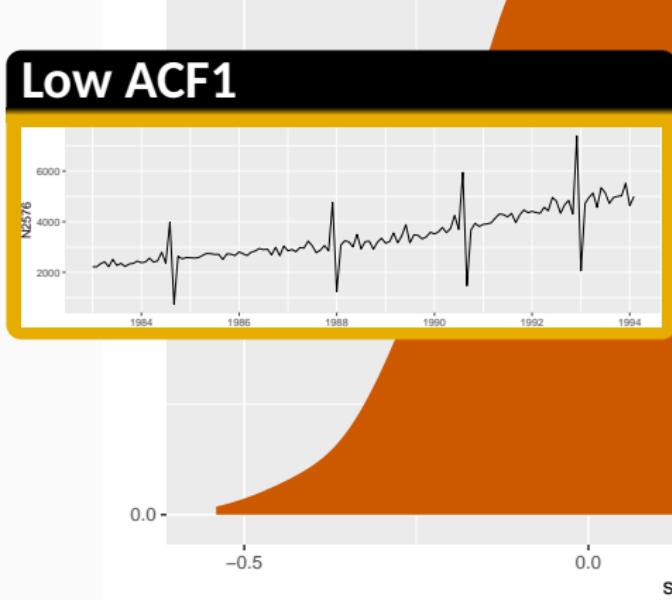
# Distribution of Trend for M3



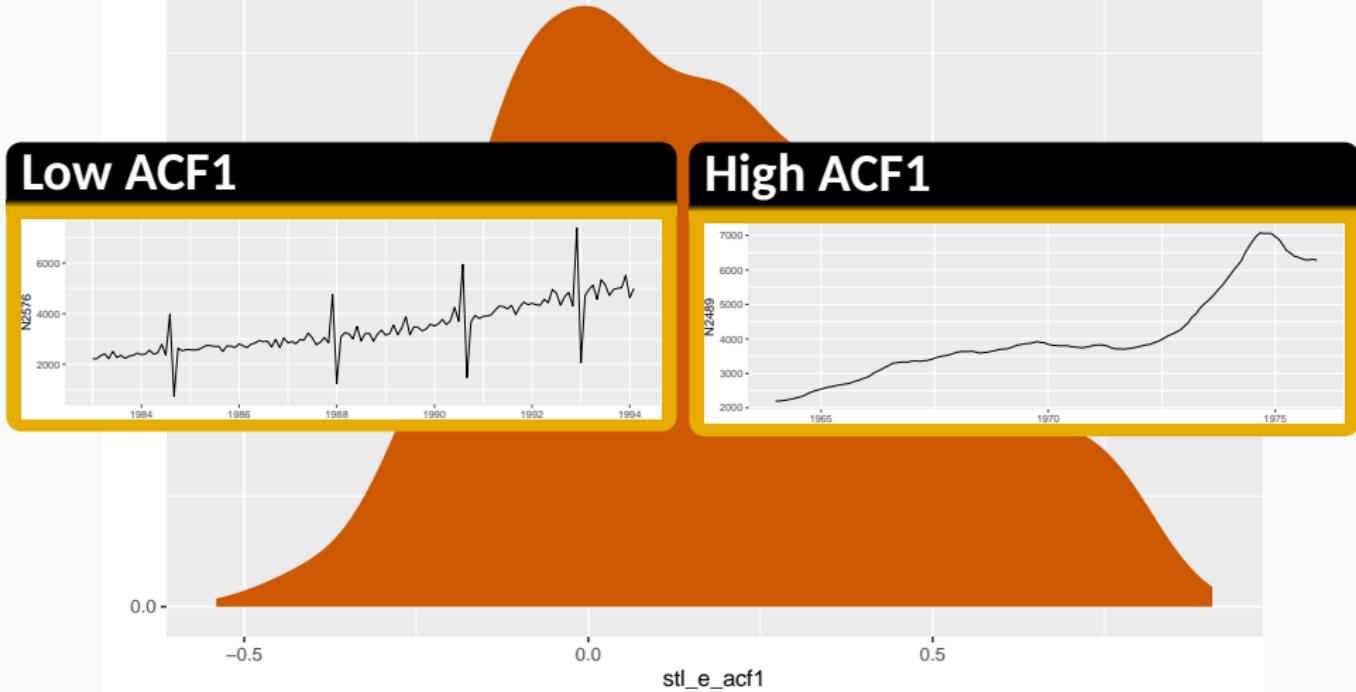
# Distribution of Residual ACF1 for M3



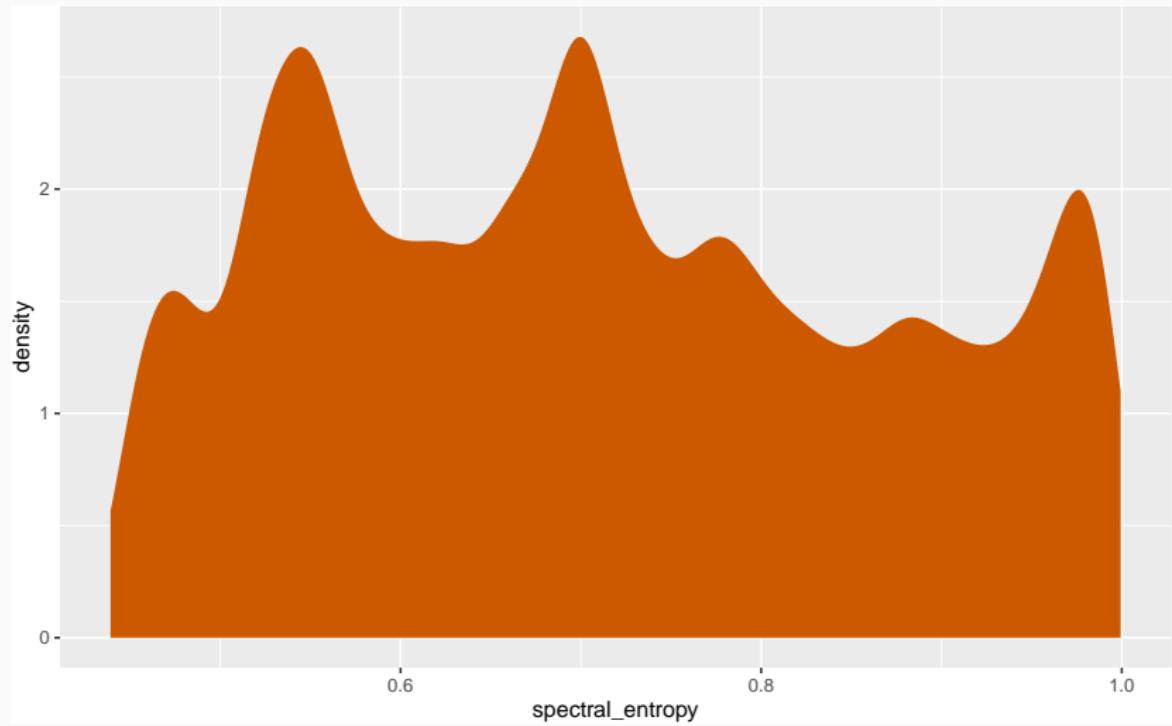
# Distribution of Residual ACF1 for M3



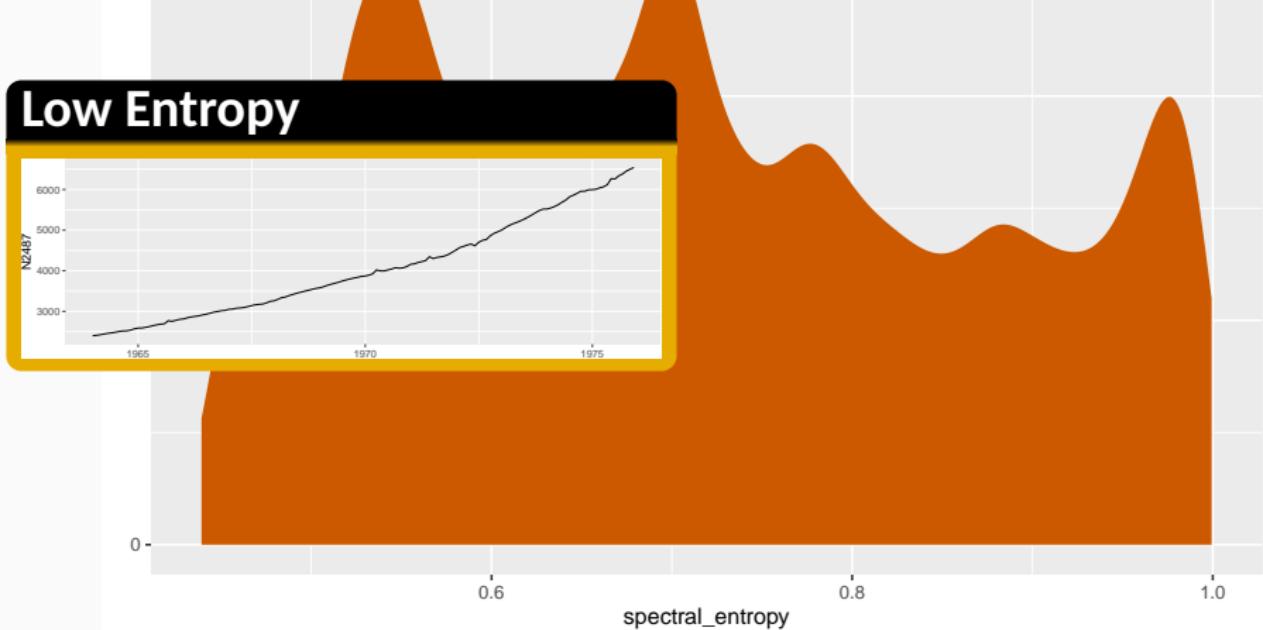
# Distribution of Residual ACF1 for M3



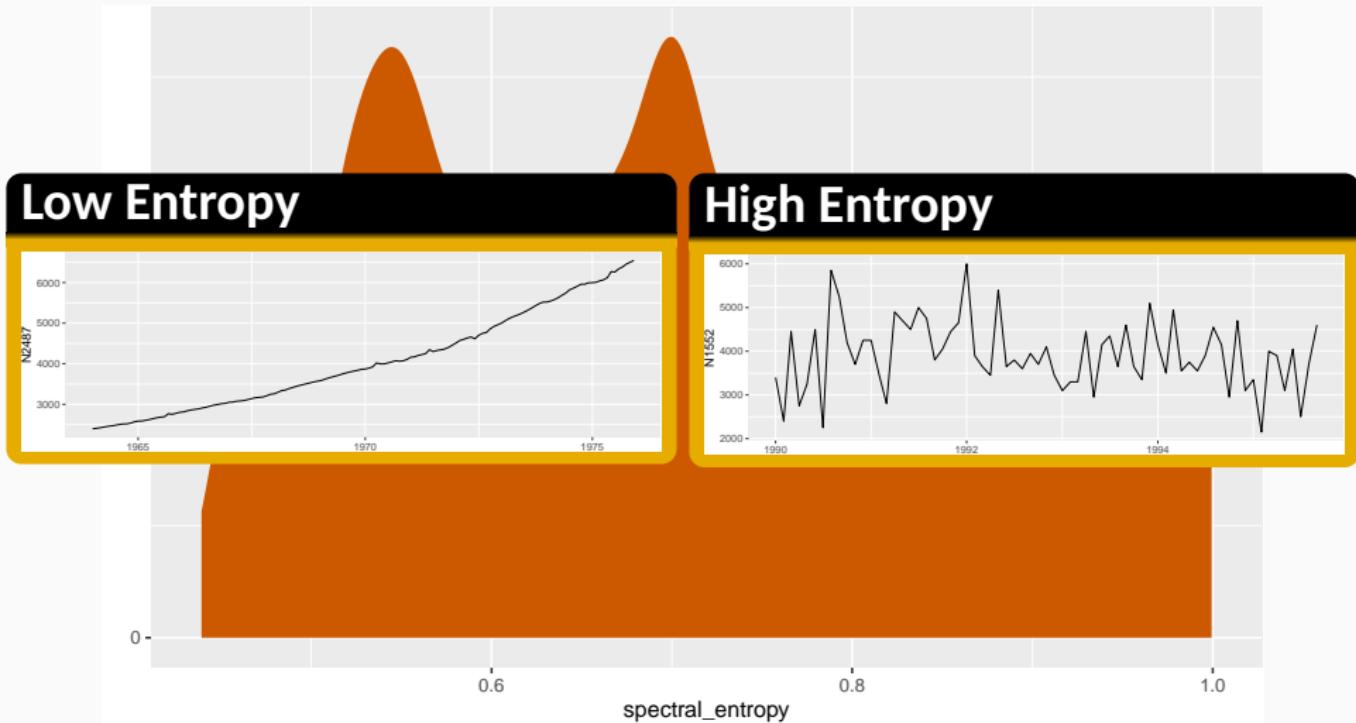
# Distribution of Spectral Entropy for M3



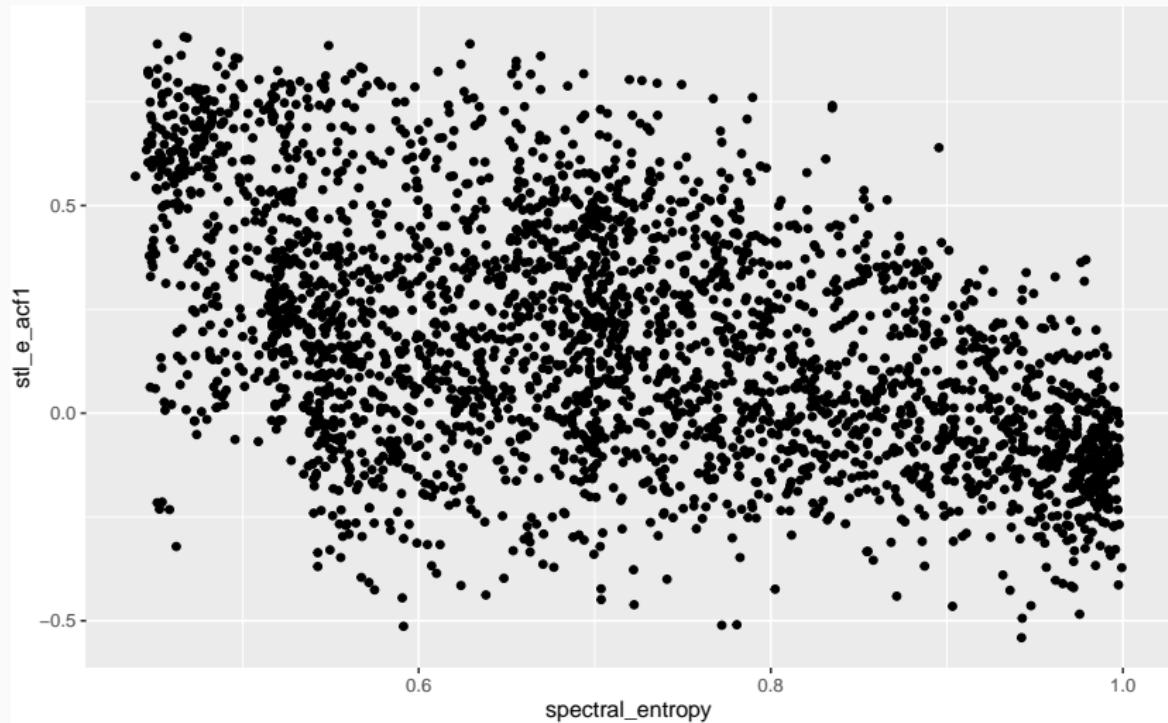
# Distribution of Spectral Entropy for M3



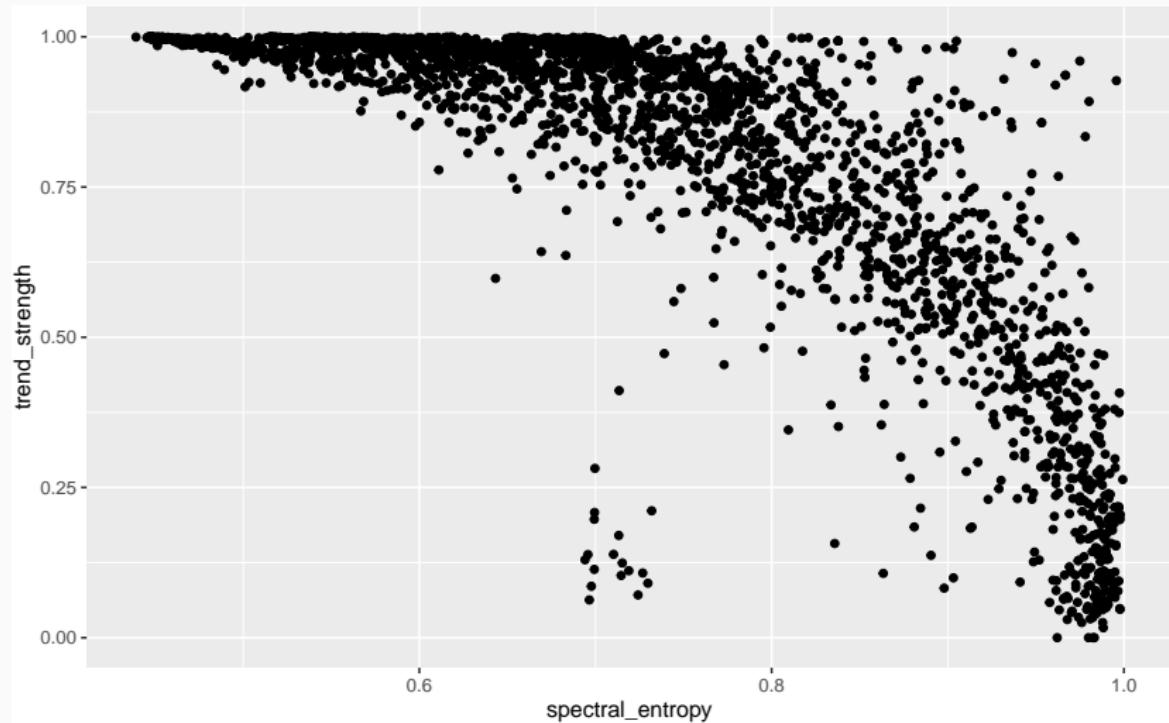
# Distribution of Spectral Entropy for M3



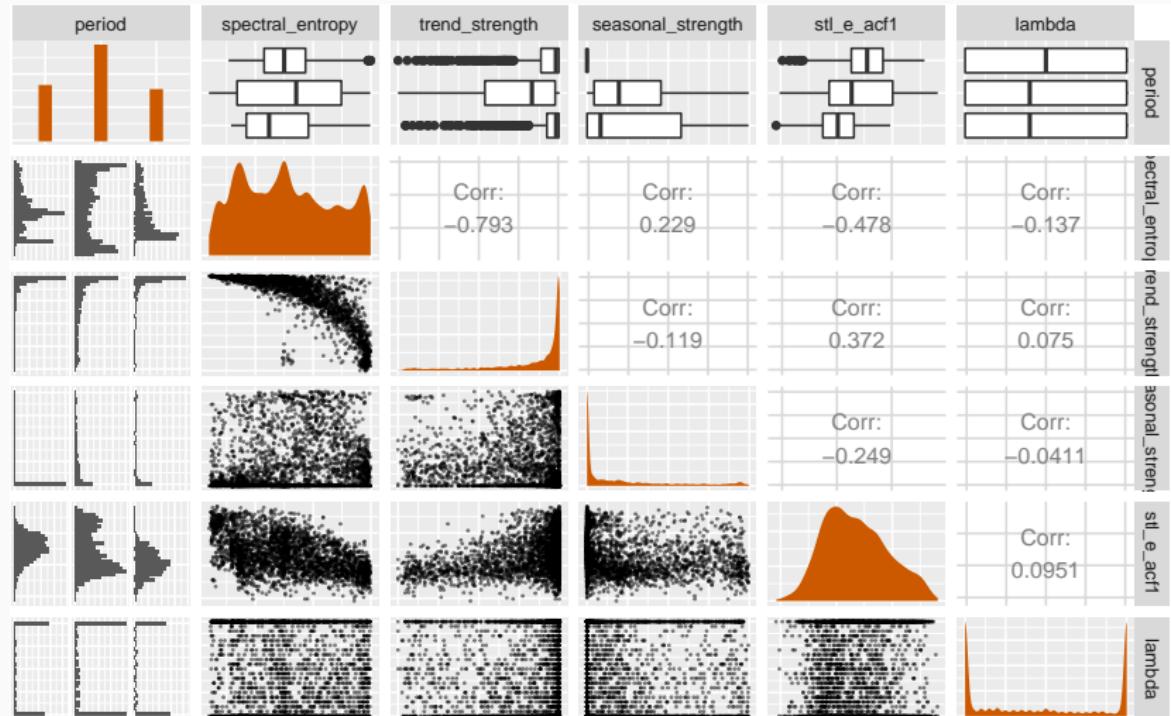
# Feature distributions



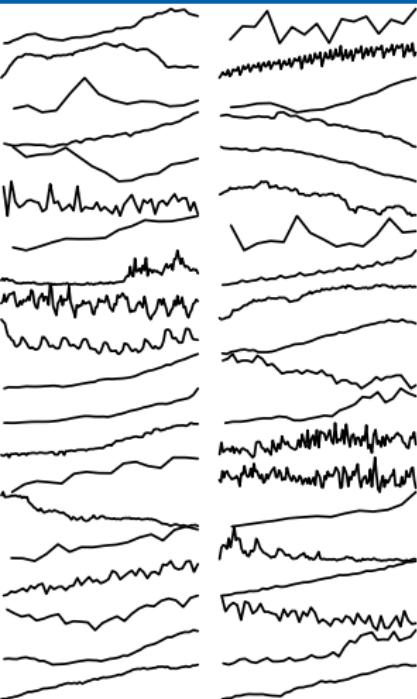
# Feature distributions



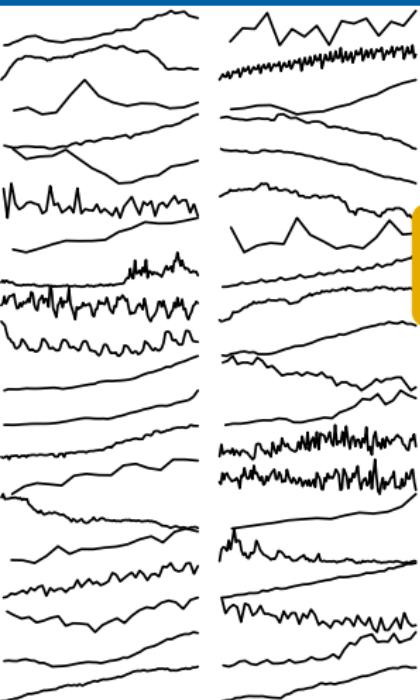
# Feature distributions



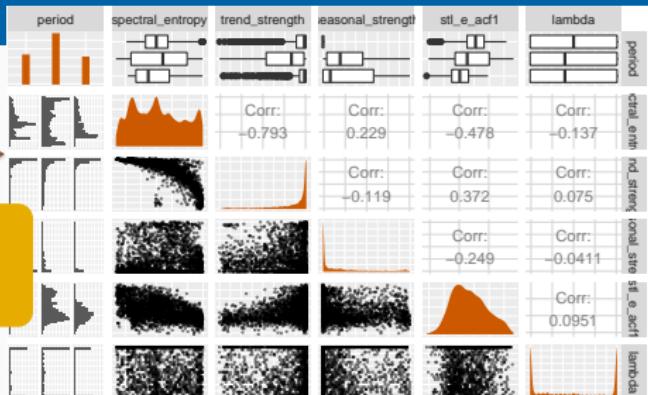
# Dimension reduction for time series



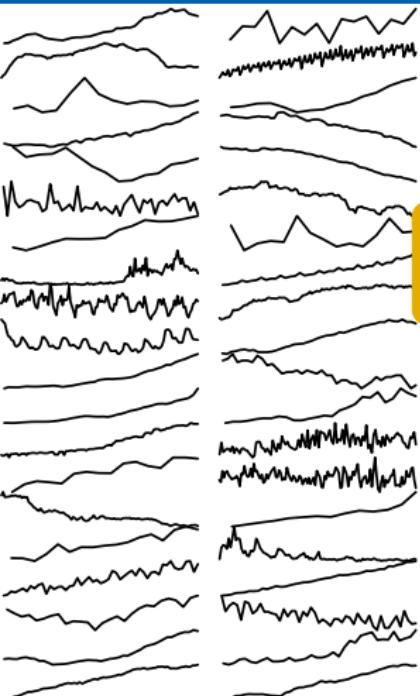
# Dimension reduction for time series



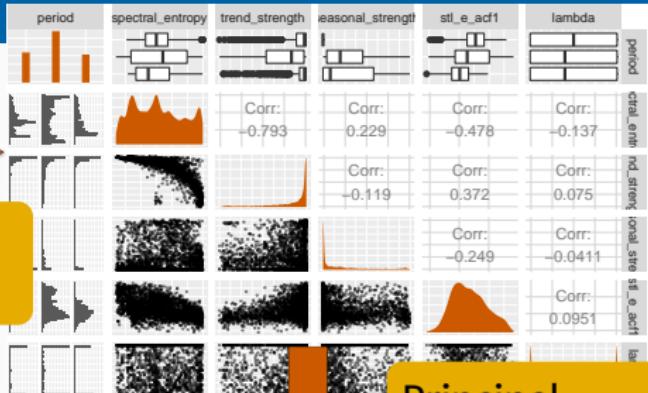
Feature  
calculation



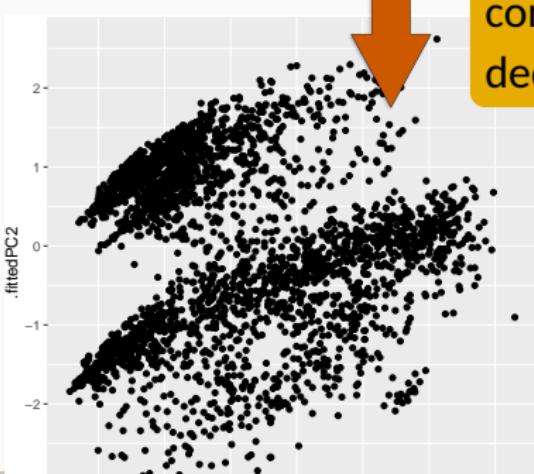
# Dimension reduction for time series



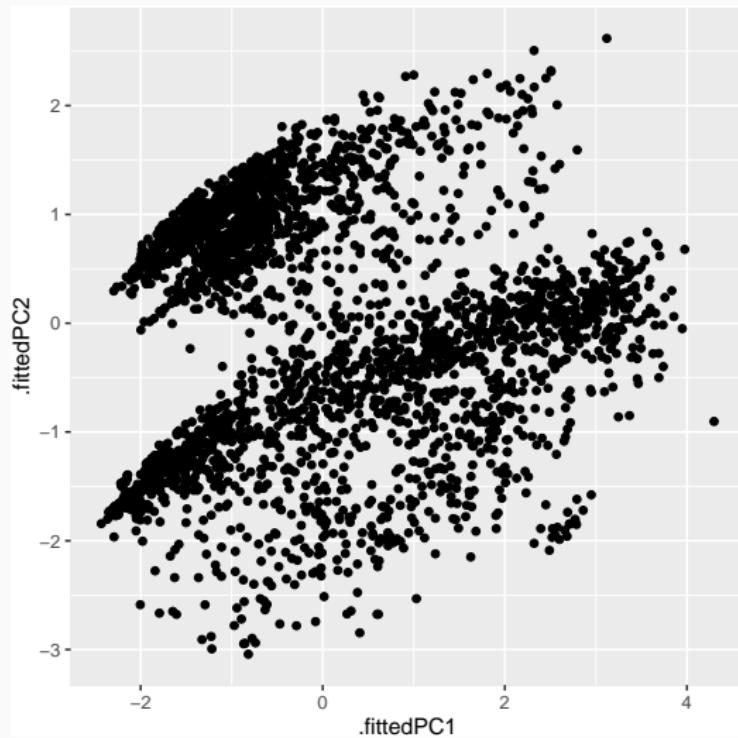
Feature  
calculation



Principal  
component  
decomposition

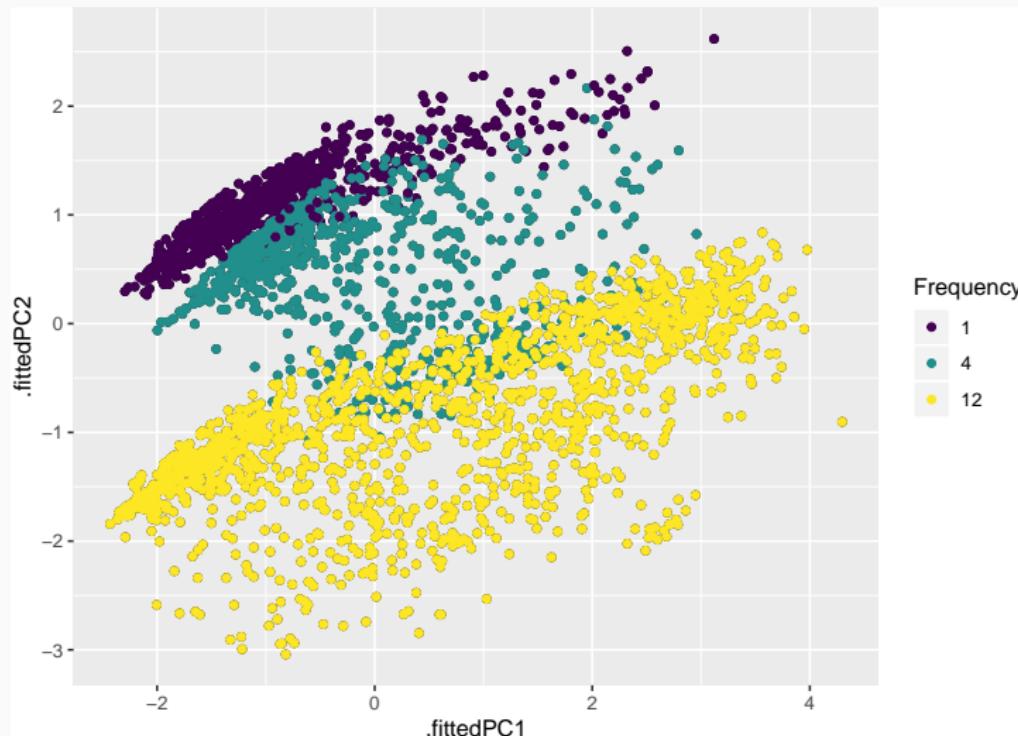


# M3 feature space

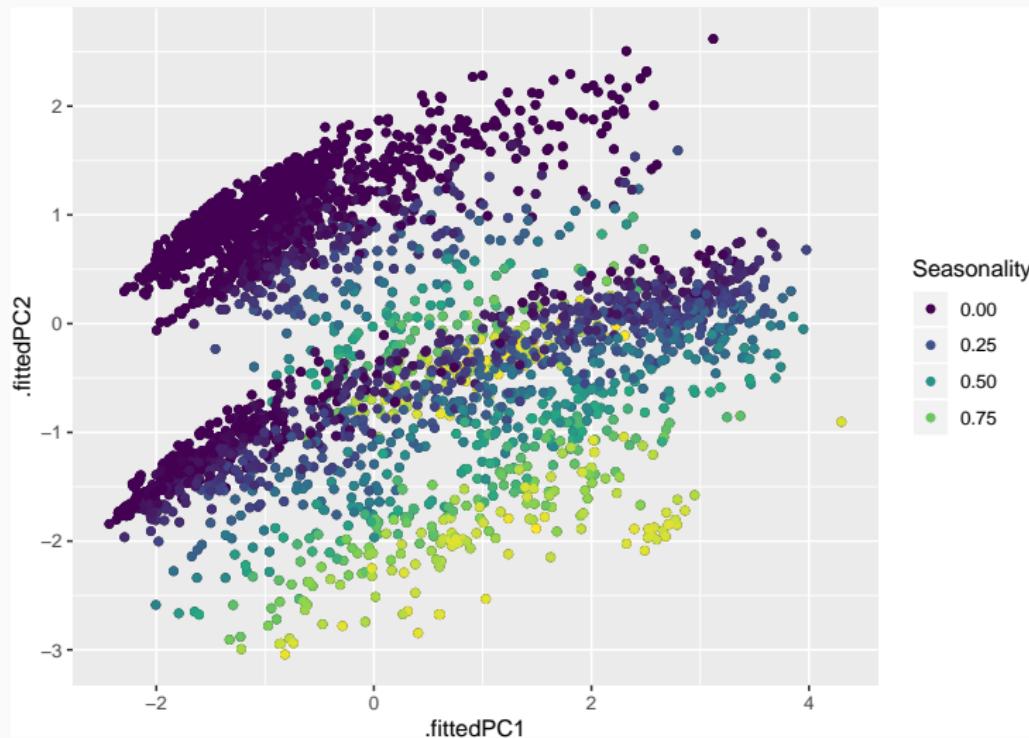


First two PCs explain  
58.5% of the variance.

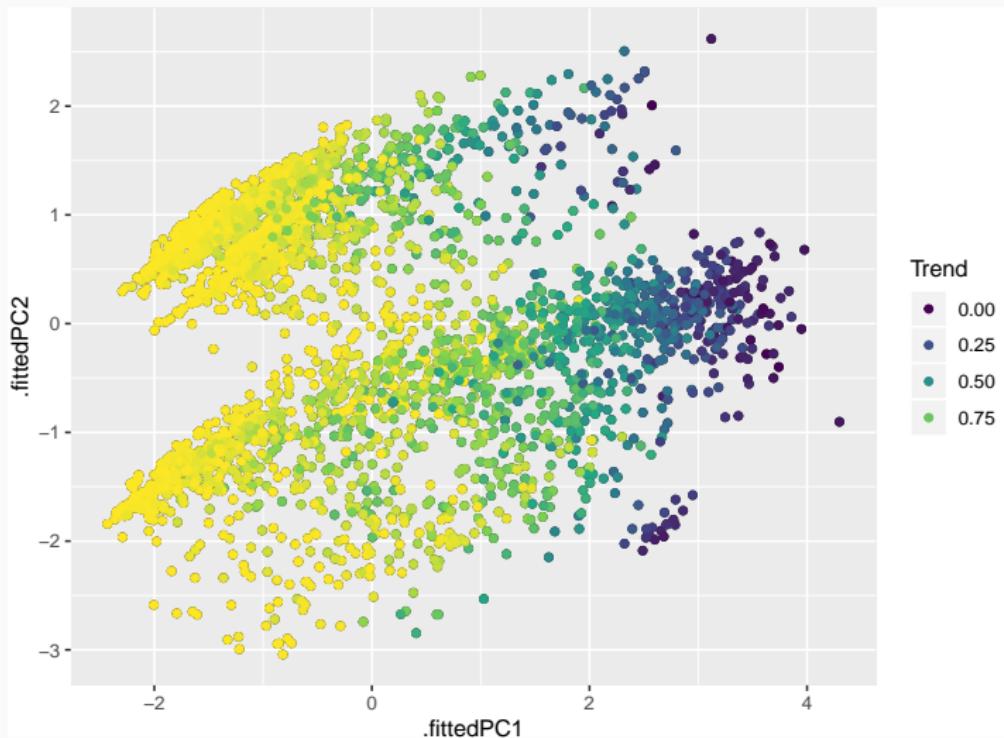
# M3 feature space



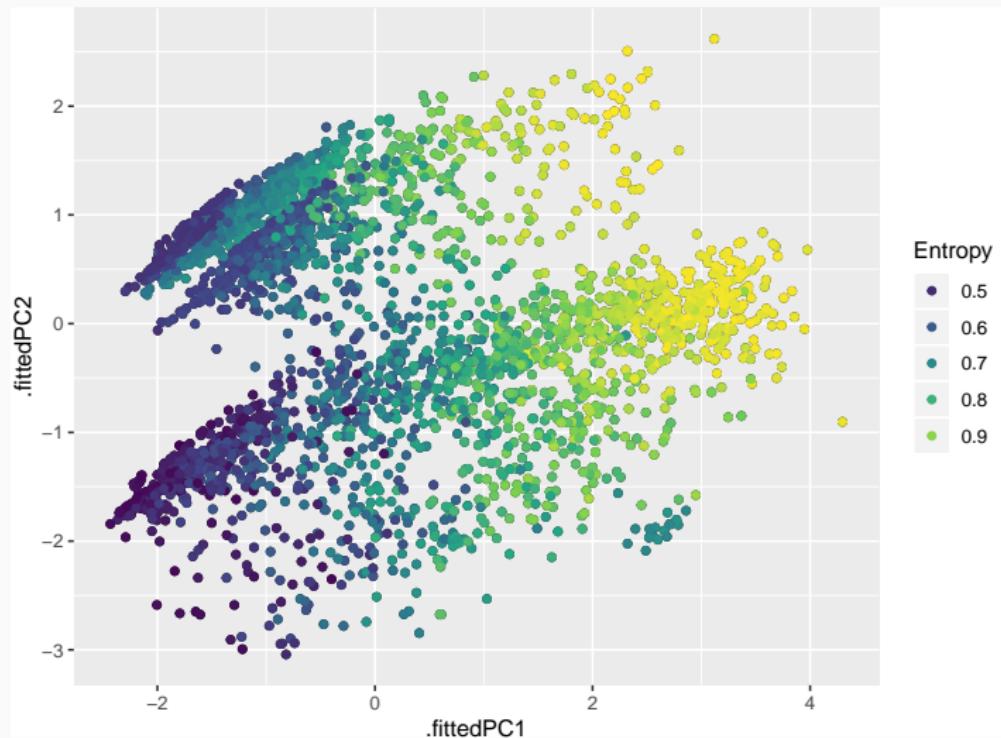
# M3 feature space



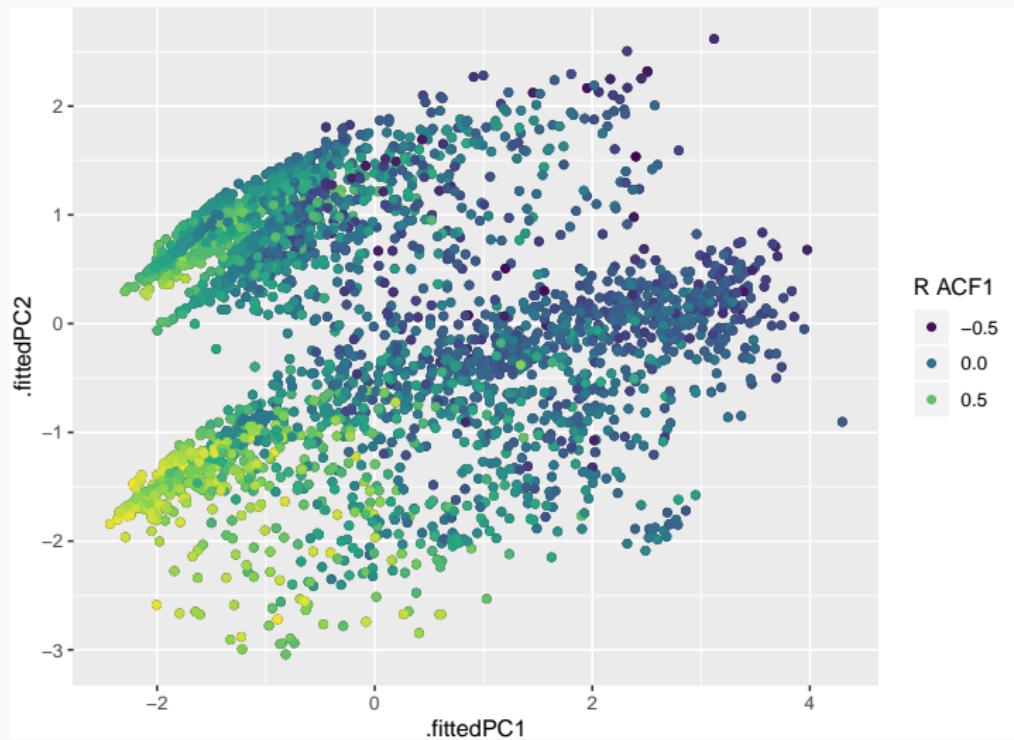
# M3 feature space



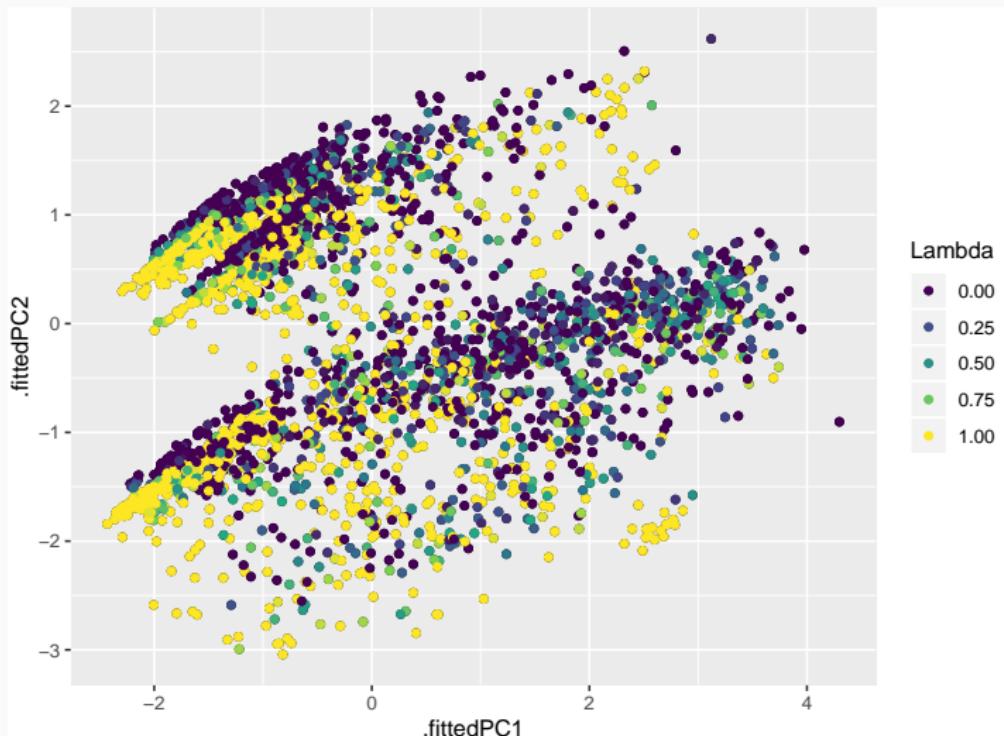
# M3 feature space



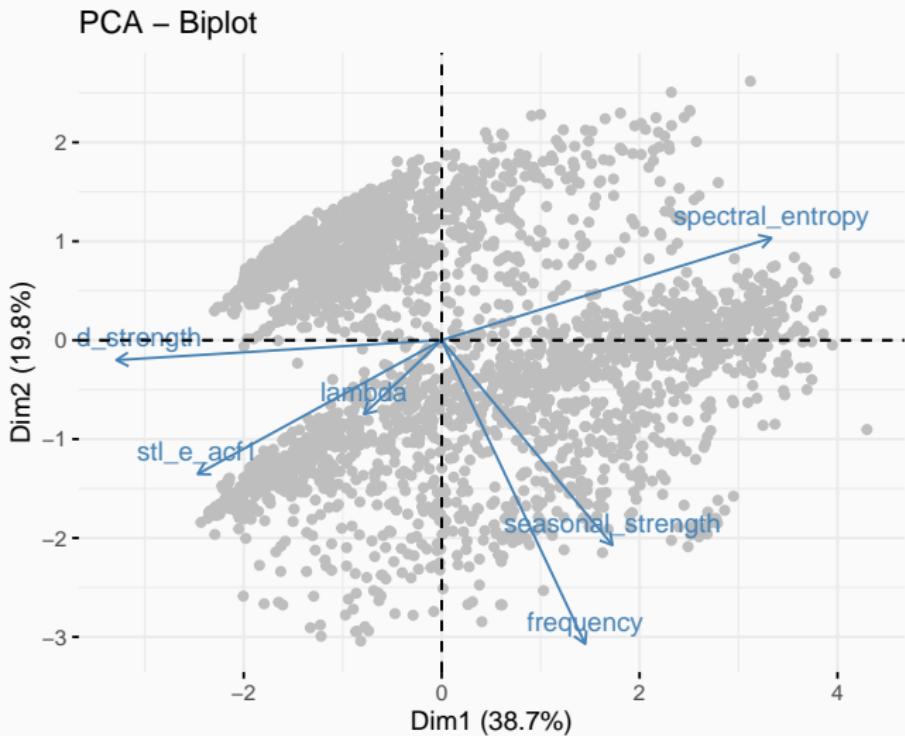
# M3 feature space



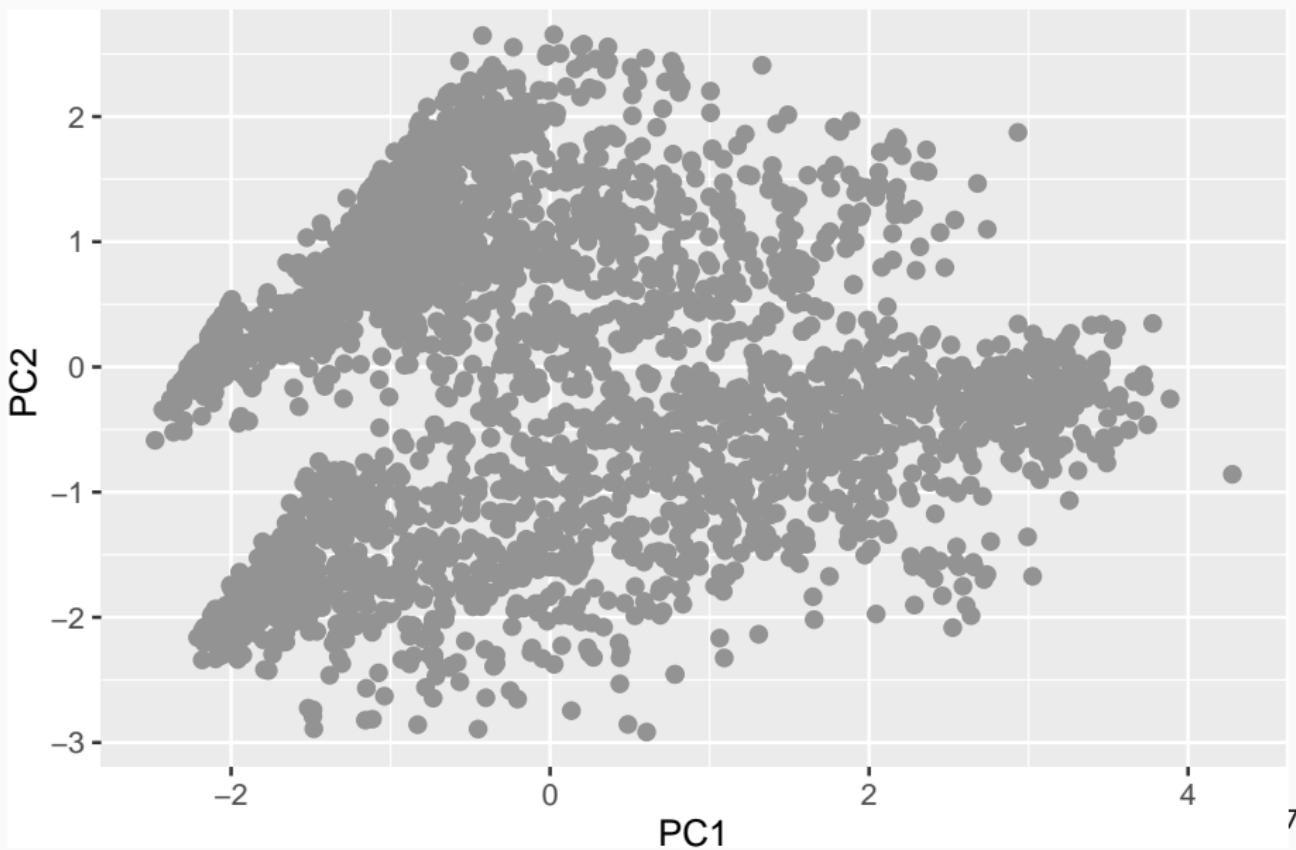
# M3 feature space



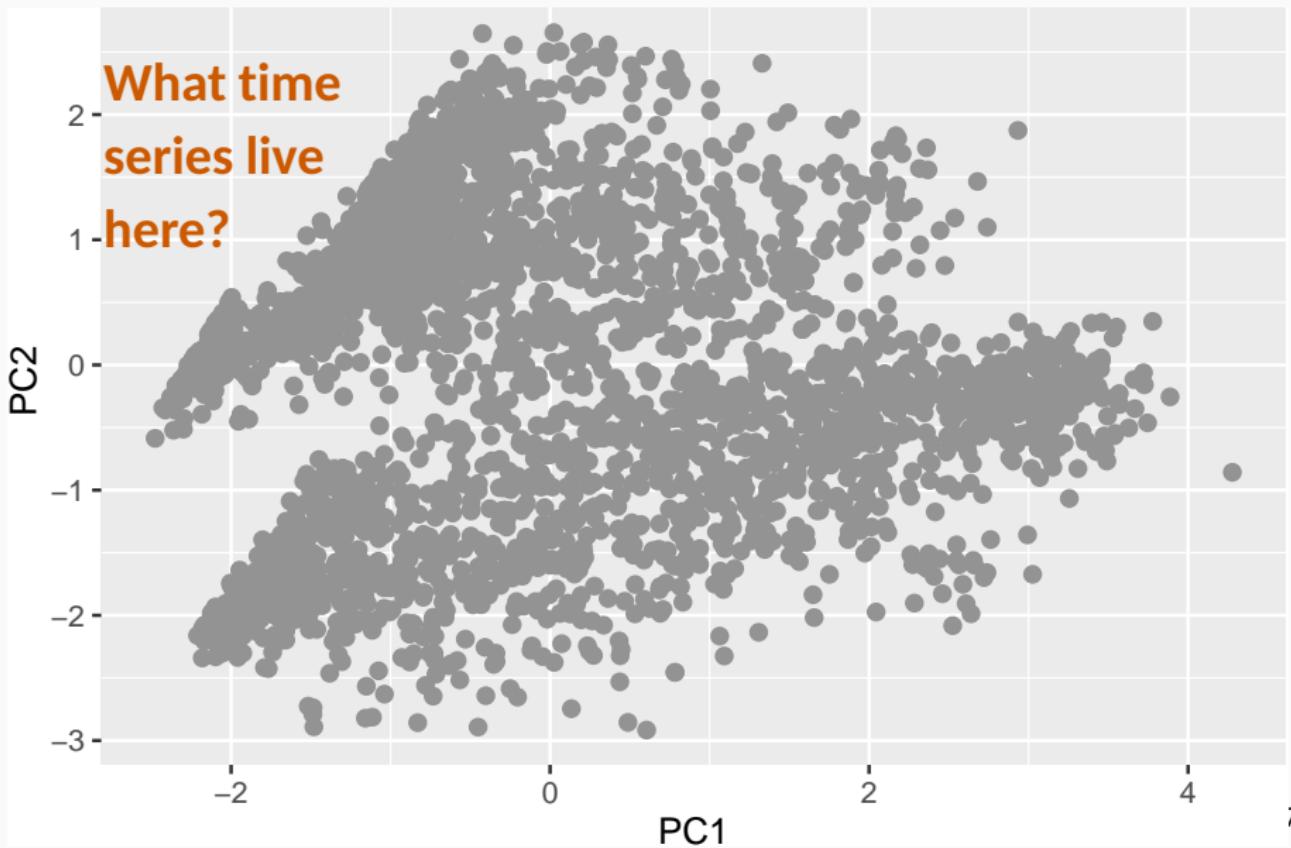
# M3 feature space



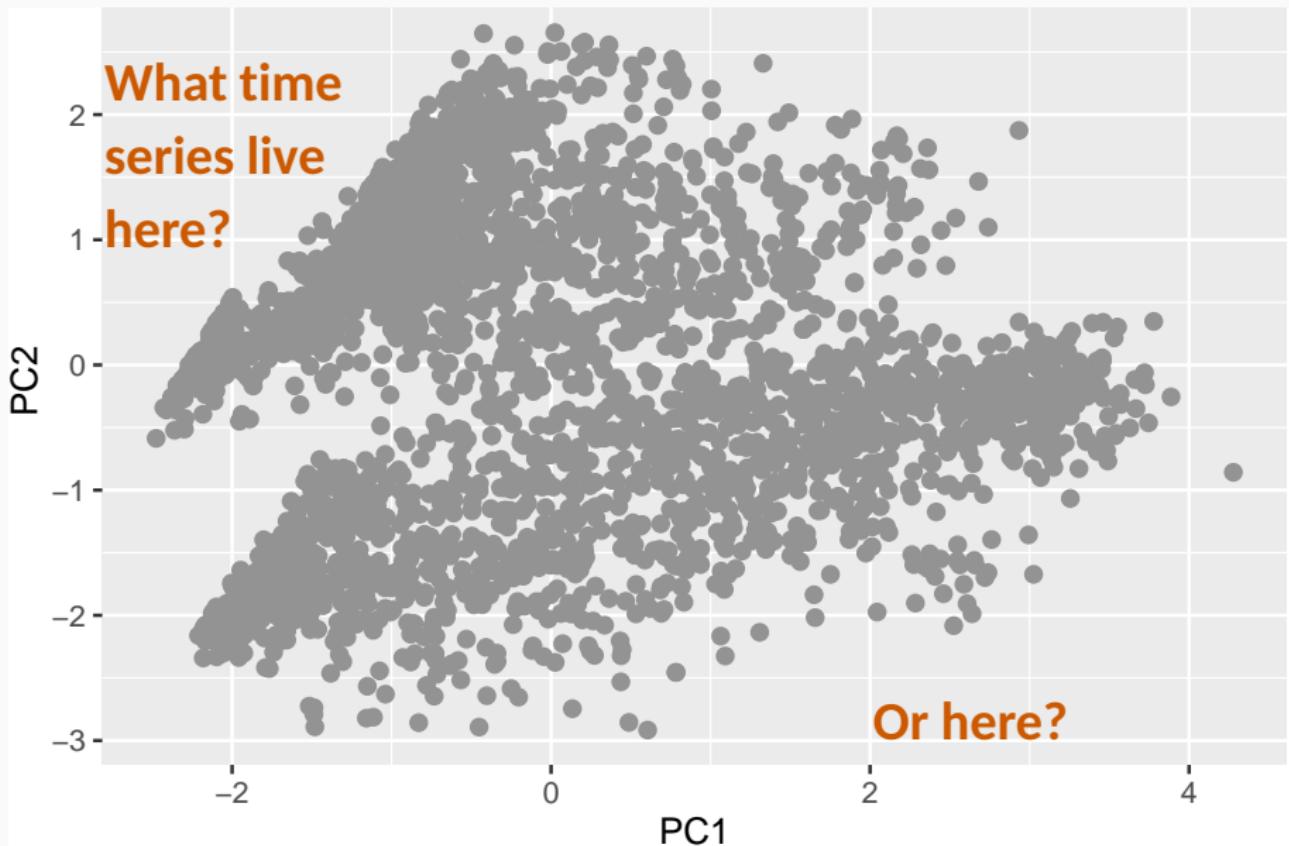
# What about the holes?



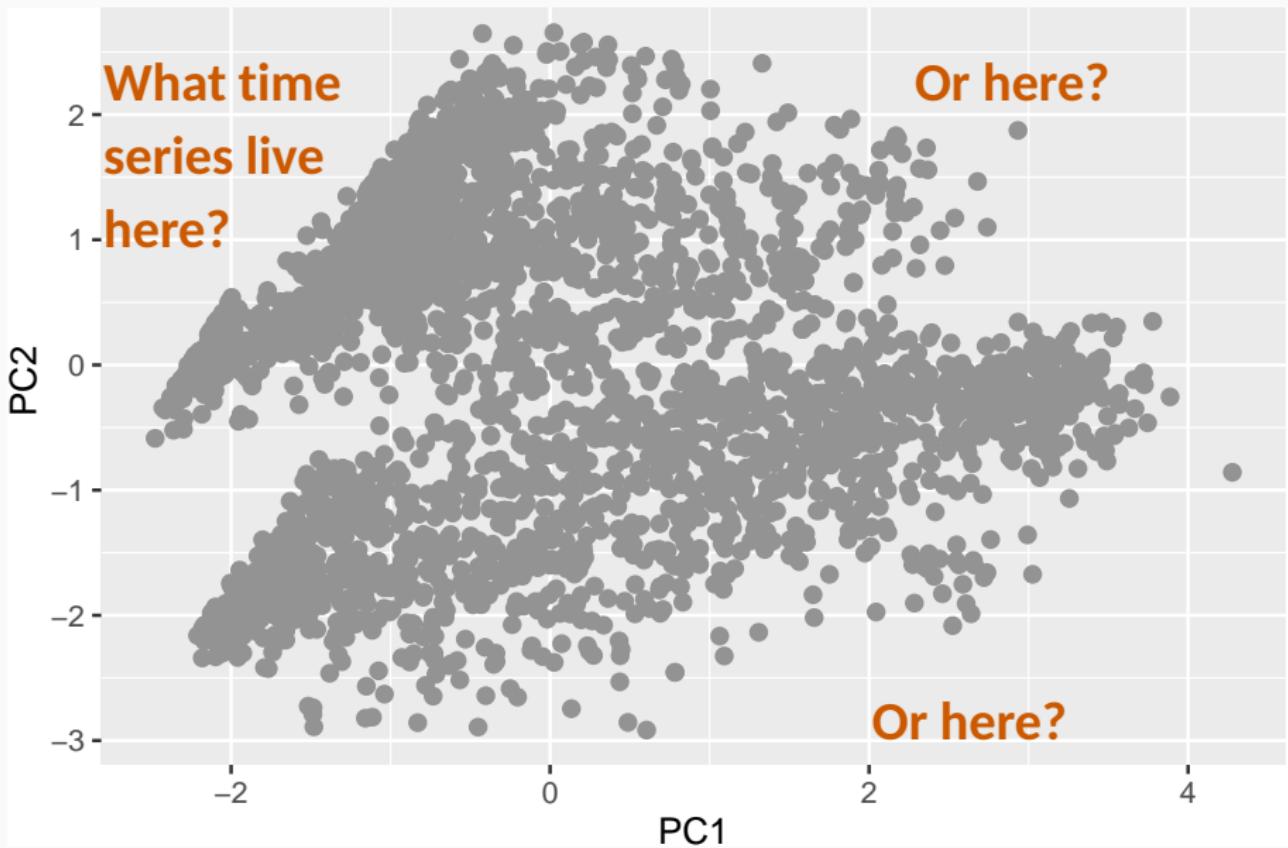
# What about the holes?



# What about the holes?



# What about the holes?



# Generating new time series

We can use the feature space to:

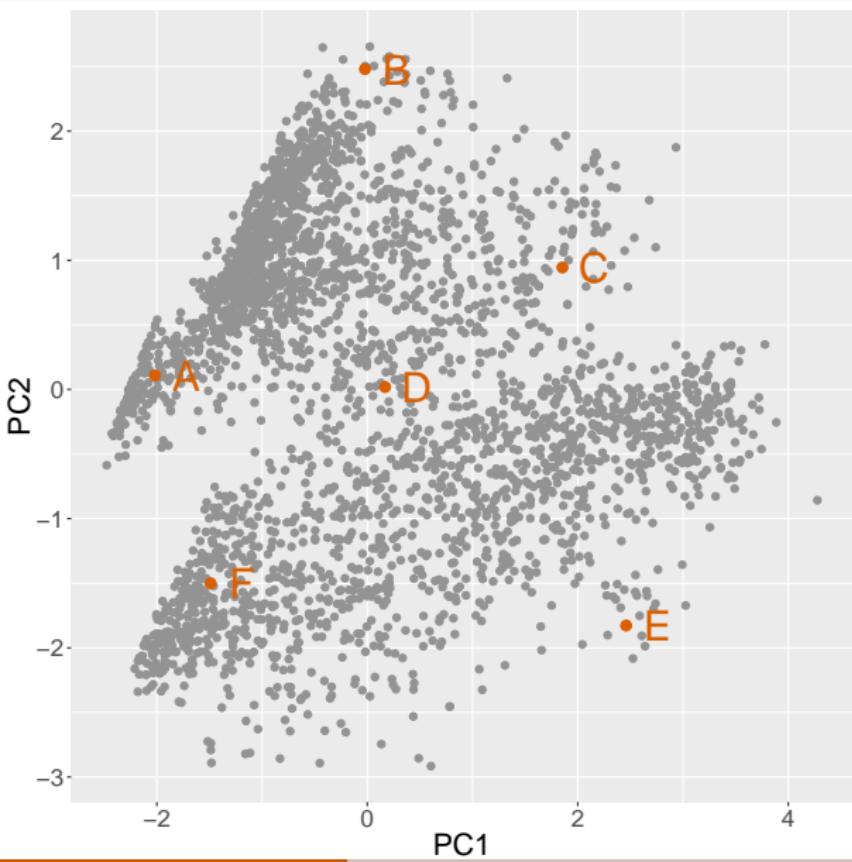
- Generate new time series with similar features to existing series
- Generate new time series where there are “holes” in the feature space.

# Generating new time series

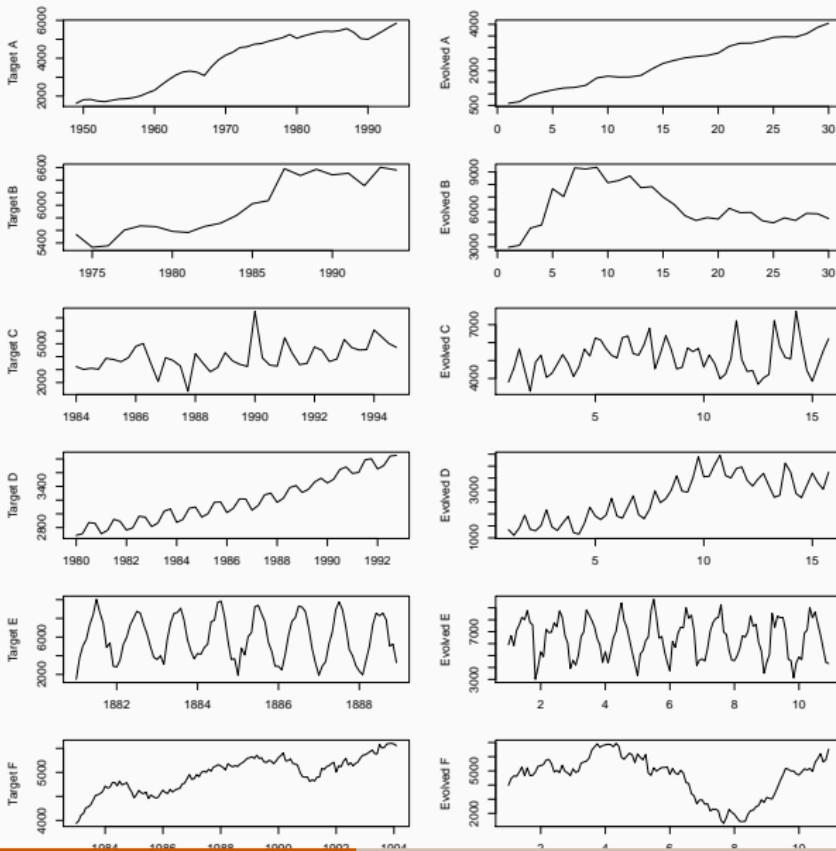
We can use the feature space to:

- 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  $T_i$ .

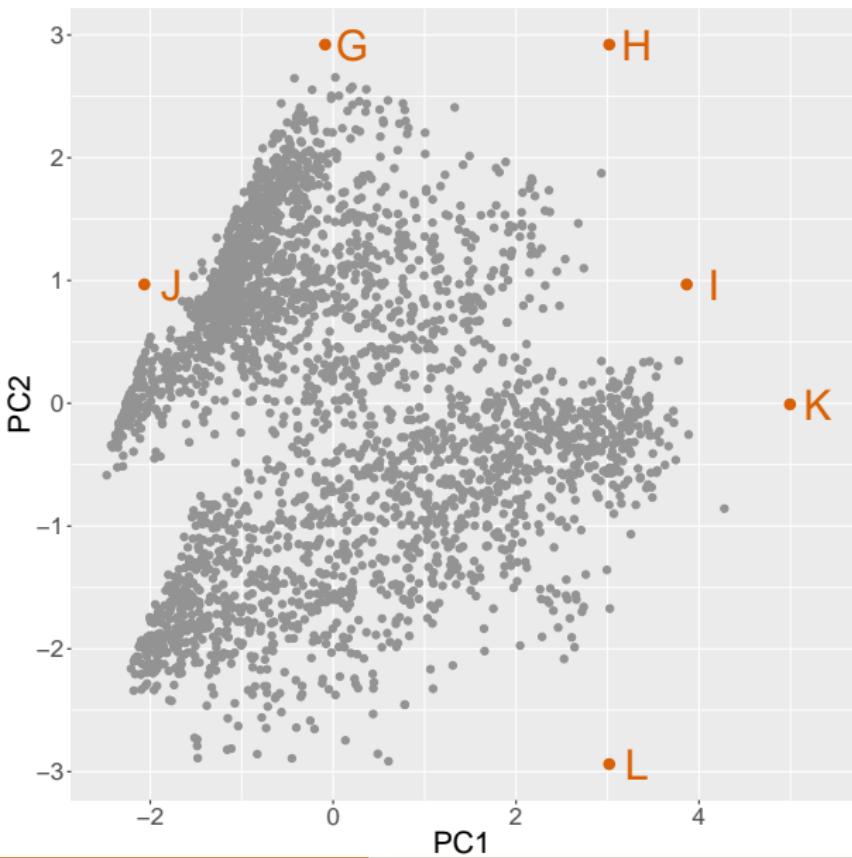
# Evolving new time series



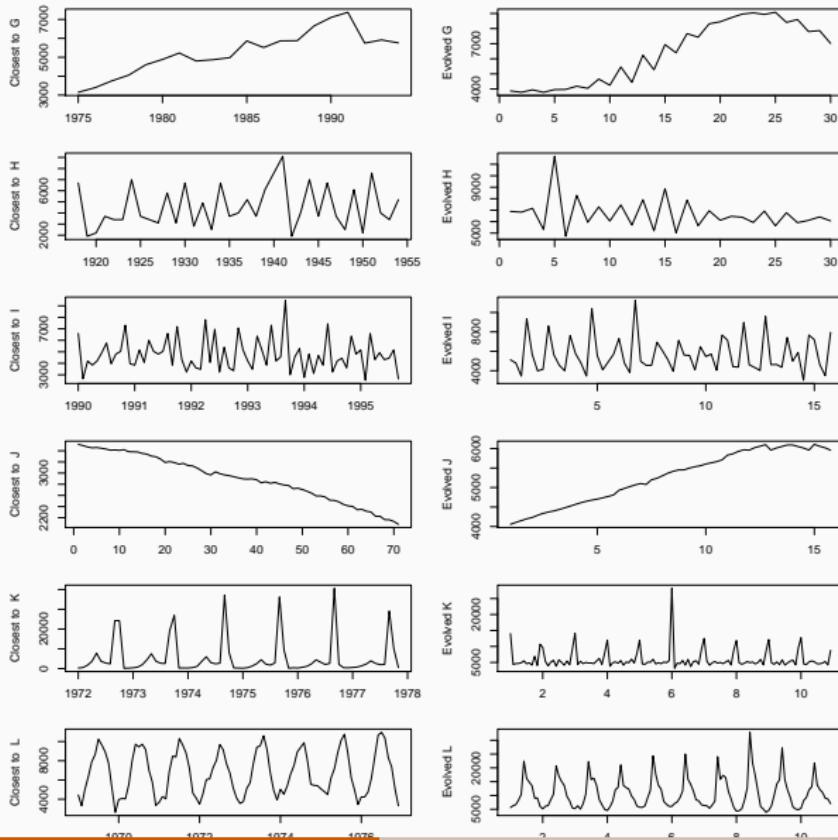
# Evolving new time series



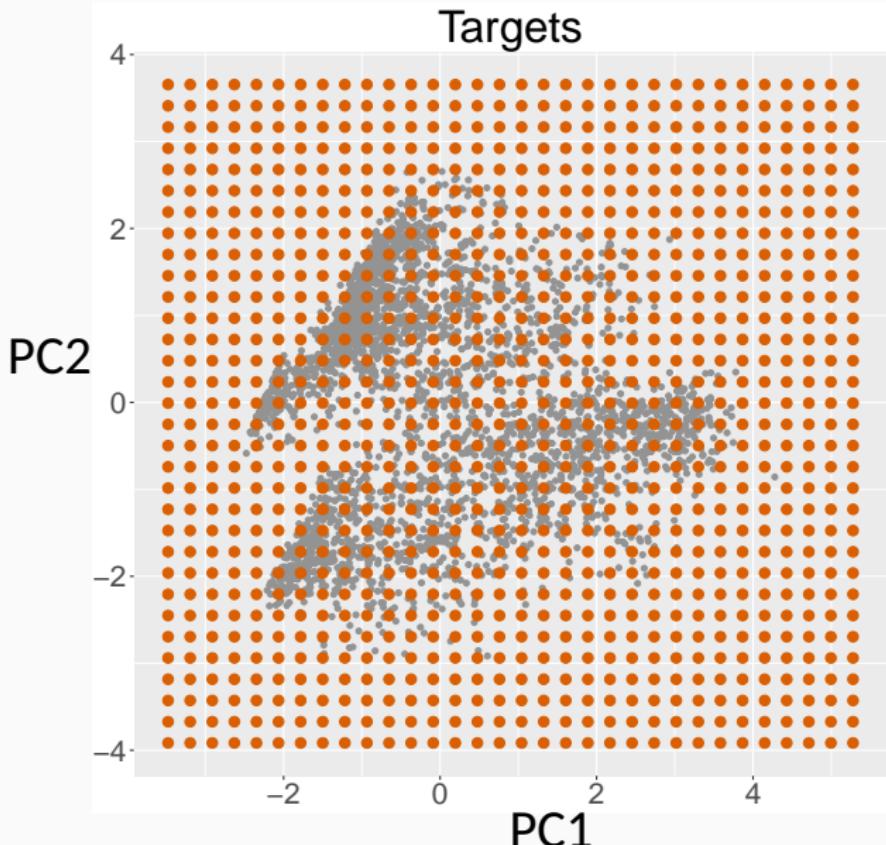
# Evolving new time series



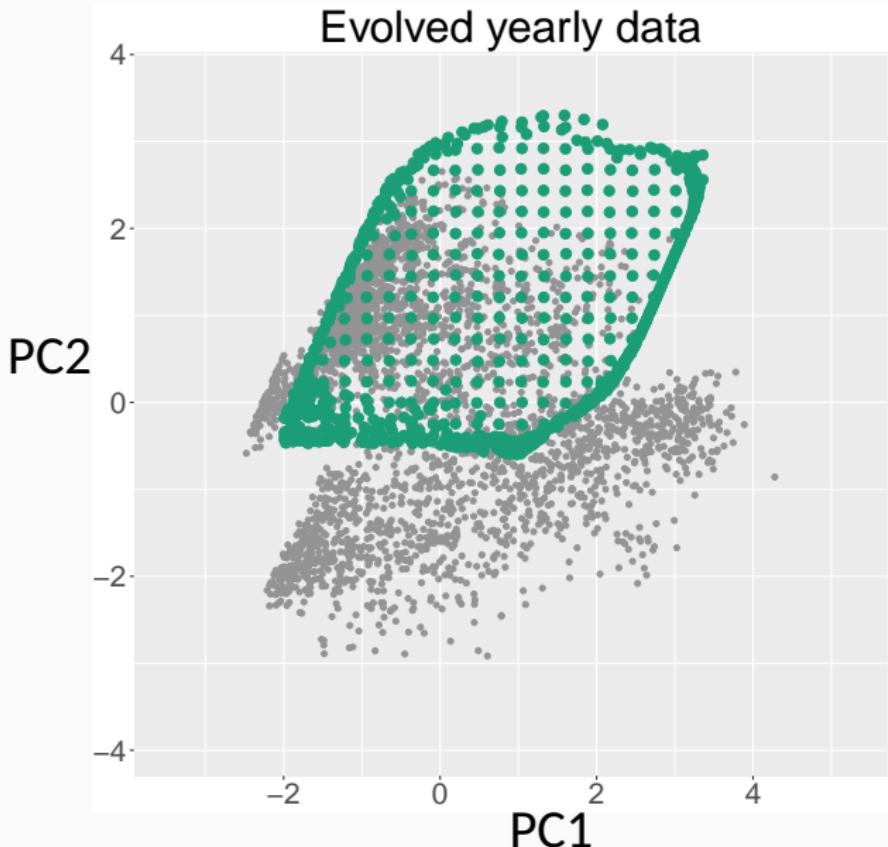
# Evolving new time series



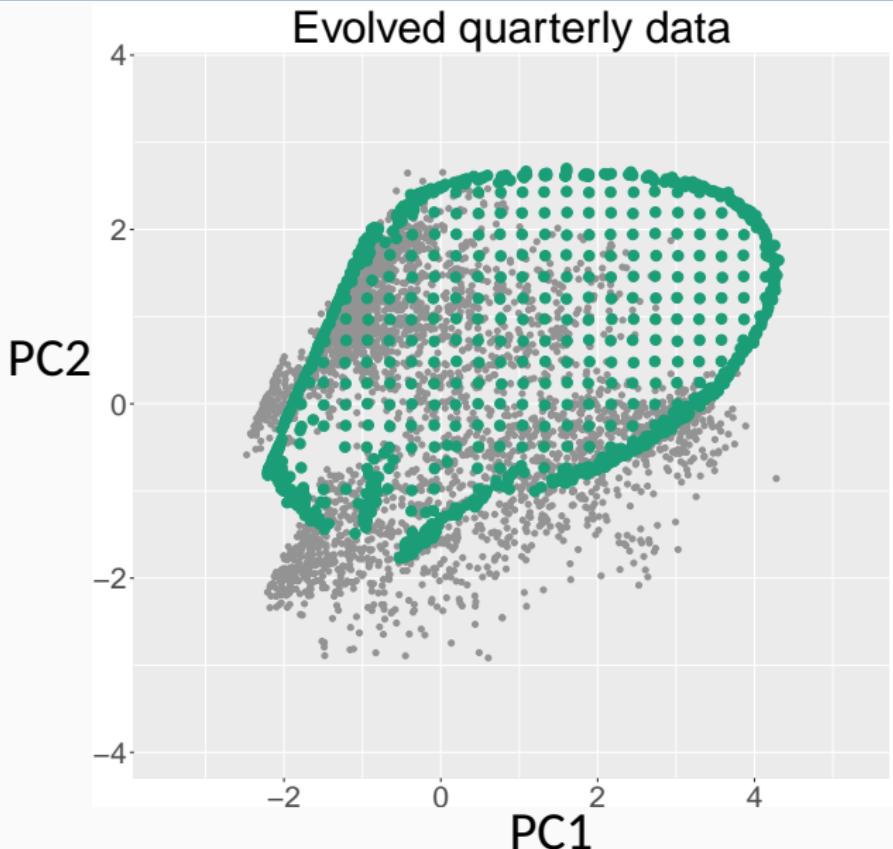
# Evolving new time series



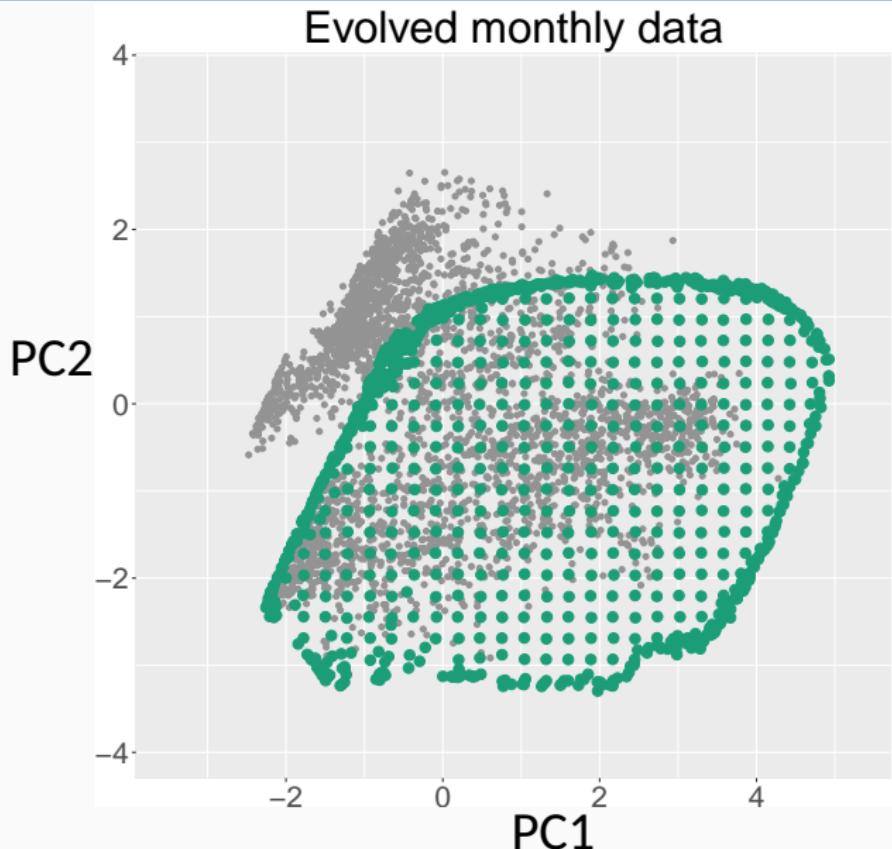
# Evolving new time series



# Evolving new time series



# Evolving new time series



# Outline

1 Visualization

2 R packages

3 Anomaly detection

4 Forecast model selection

5 Forecast model averaging



# 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.

# Holidays by state

```
holidays <- tourism %>%  
  filter(Purpose=="Holiday") %>%  
  group_by(State) %>%  
  summarise(Trips = sum(Trips))
```

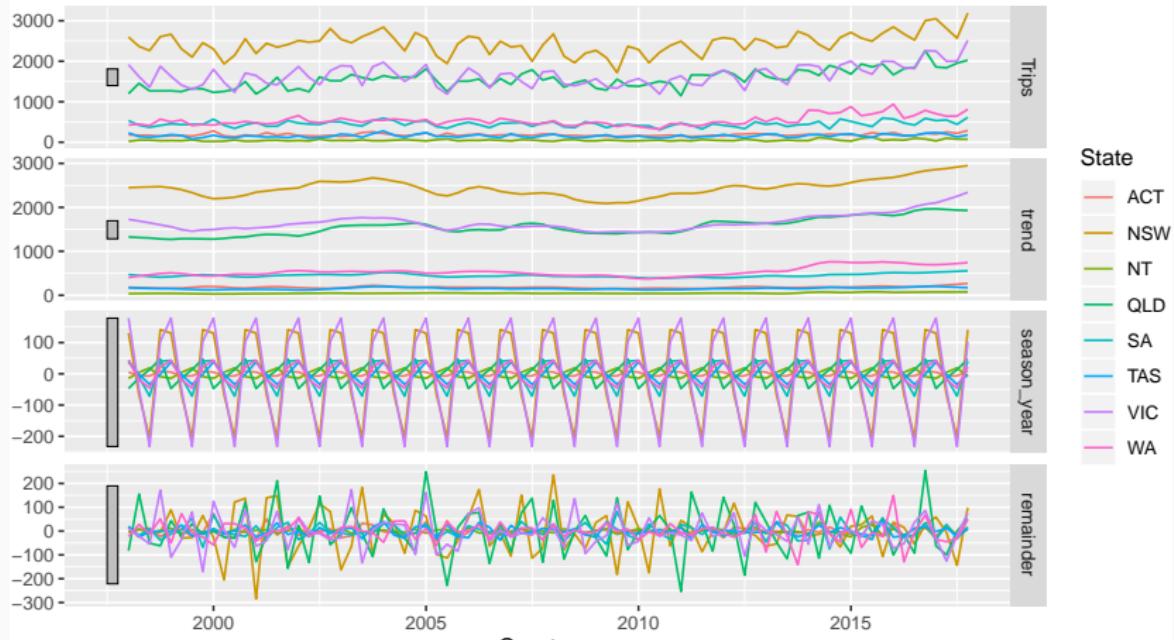
```
## # A tsibble: 640 x 3 [1Q]  
## # Key:      State [8]  
##   Quarter State Trips  
##       <qtr> <chr> <dbl>  
## 1 1998 Q1 ACT     183.  
## 2 1998 Q2 ACT     172.  
## 3 1998 Q3 ACT     173.  
## 4 1998 Q4 ACT     146.  
## 5 1999 Q1 ACT     162.  
## 6 1999 Q2 ACT     165.  
## 7 1999 Q3 ACT     151.  
## 8 1999 Q4 ACT     200.  
## 9 2000 Q1 ACT     279.  
## 10 2000 Q2 ACT    157.
```

# Decompositions

```
holidays %>% STL(Trips ~ season(window = "periodic")) %>%  
  autoplot()
```

STL decomposition

Trips = trend + season\_year + remainder



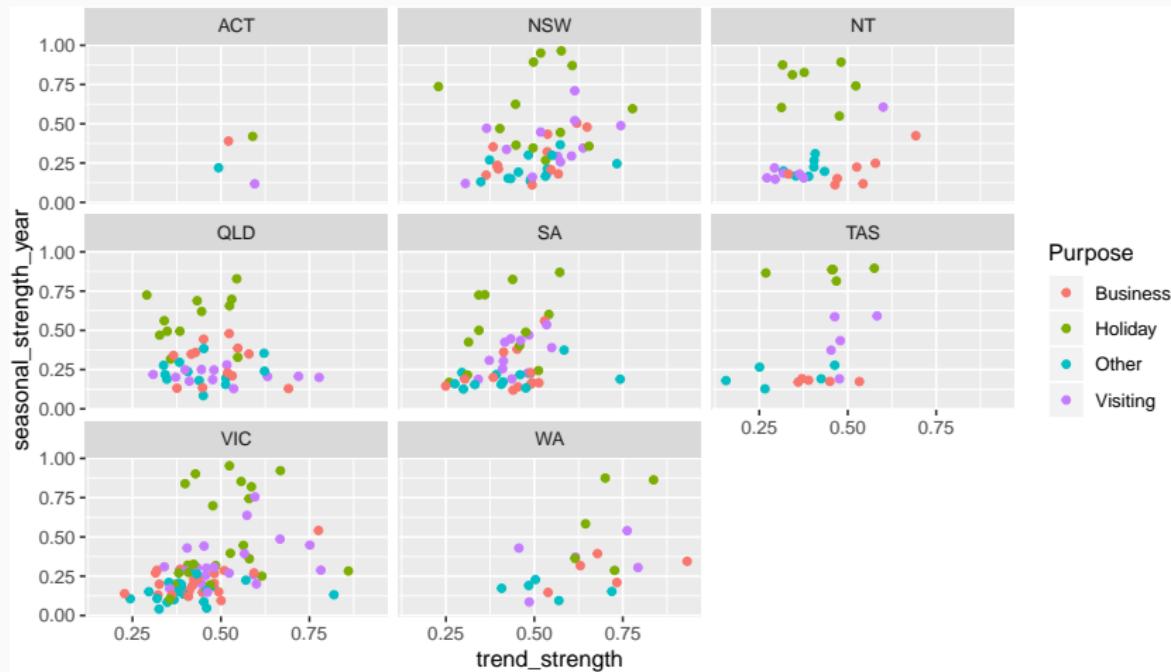
# Feature extraction and statistics

```
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>
```

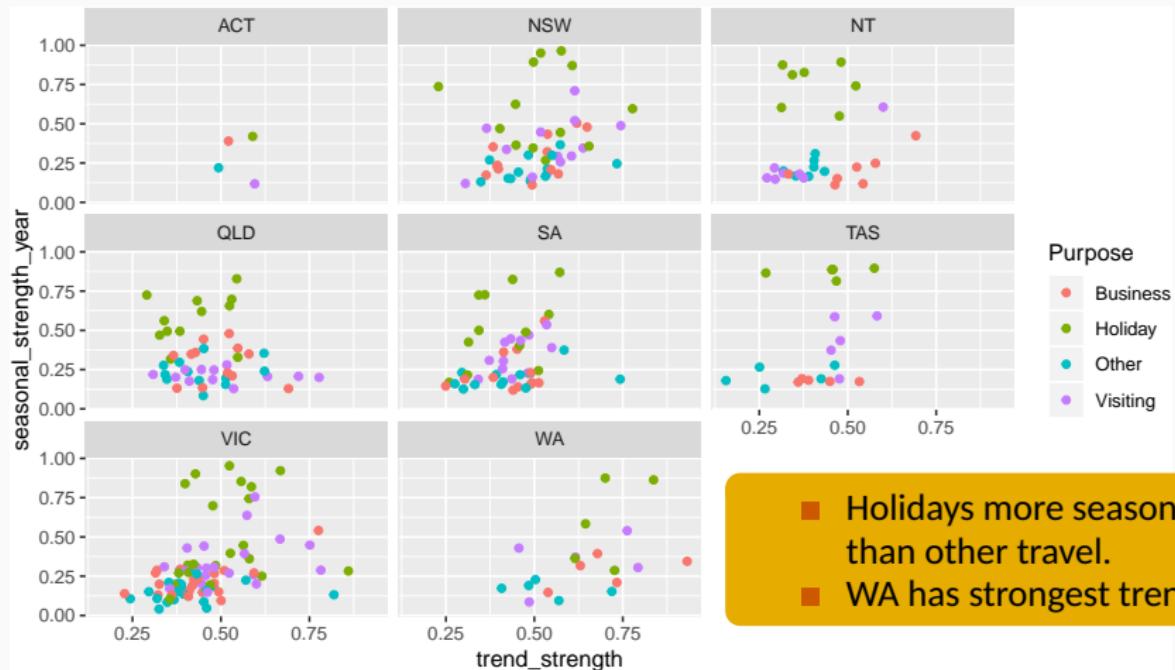
# Feature extraction and statistics

```
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:

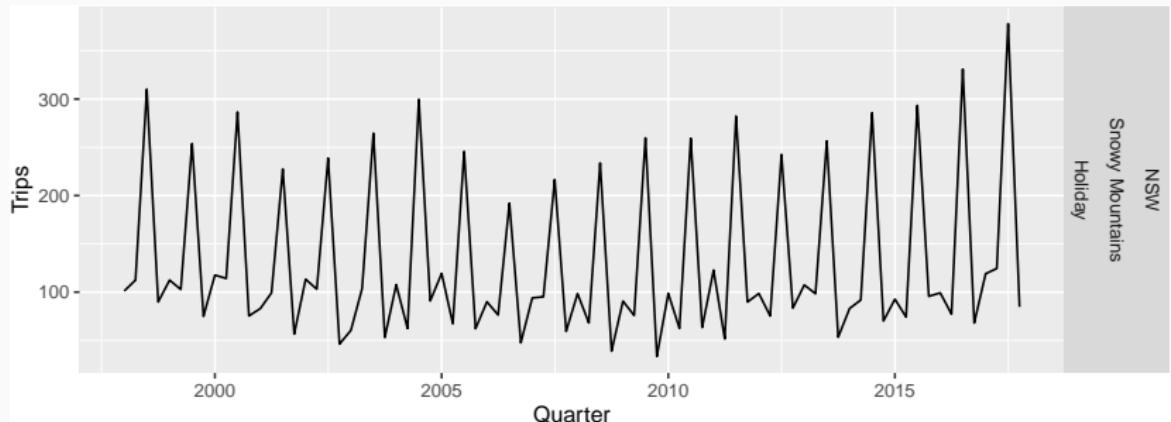
```
most_seasonal <- tourism %>%
  features(Trips, feature_set(tags="stl")) %>%
  filter(seasonal_strength_year == max(seasonal_strength_year))
```

# 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)
```

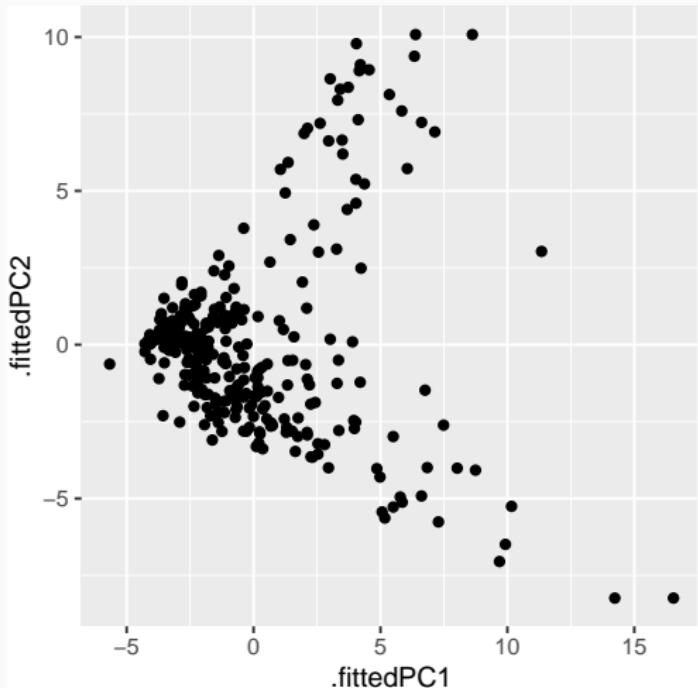
```
## # A tibble: 304 x 92  
##   .rownames Region State Purpose trend_strength  
##   <fct>     <chr>  <chr>  <chr>          <dbl>  
## 1 1          Adela~ SA     Busine~        0.451  
## 2 2          Adela~ SA     Holiday        0.541  
## 3 3          Adela~ SA     Other          0.743  
## 4 4          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>,  
## #   diff1 ~ f1 + f2; diff1 ~ f12 + f22; diff2 ~ 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)
```

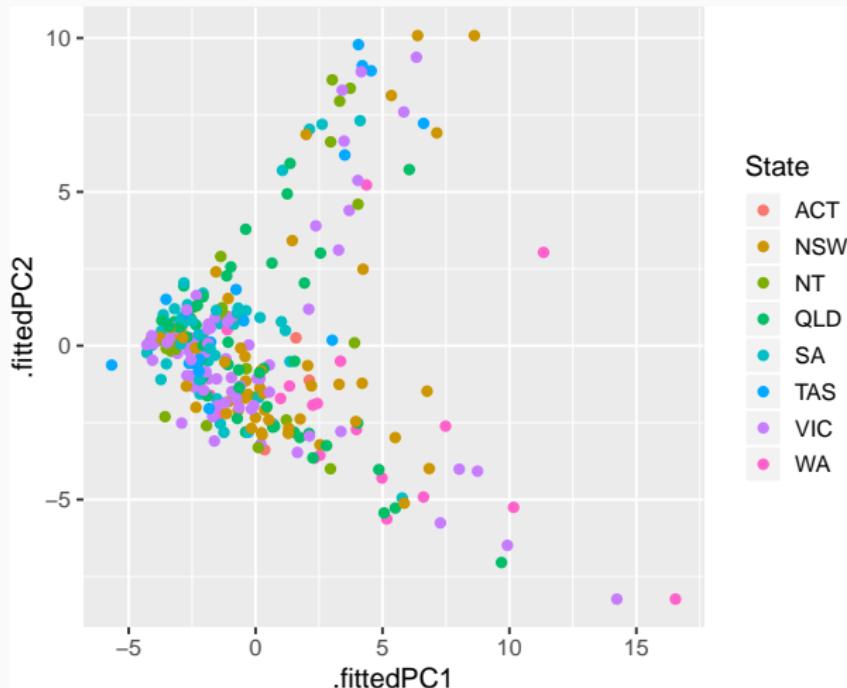
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

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

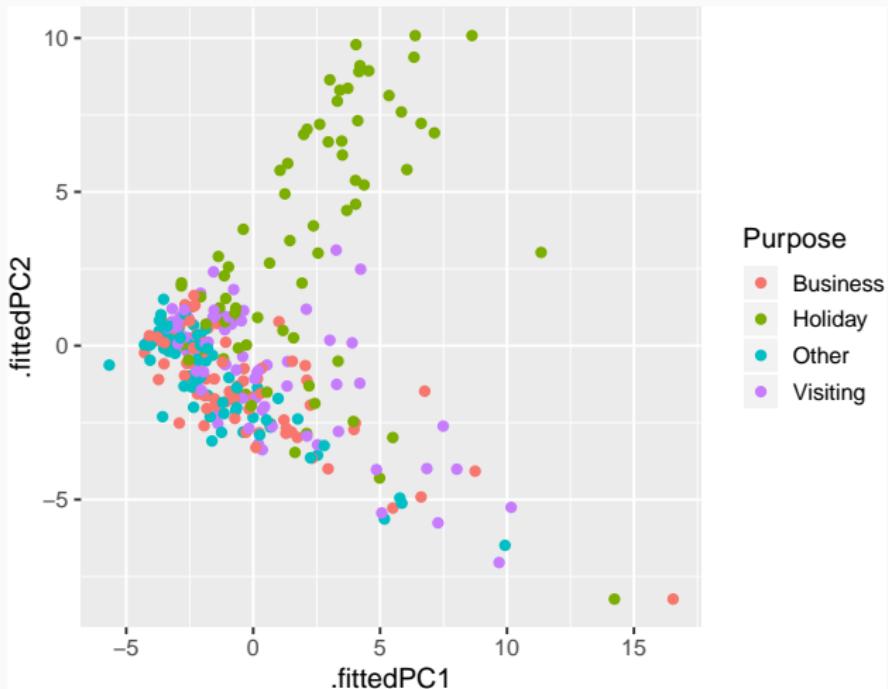
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

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

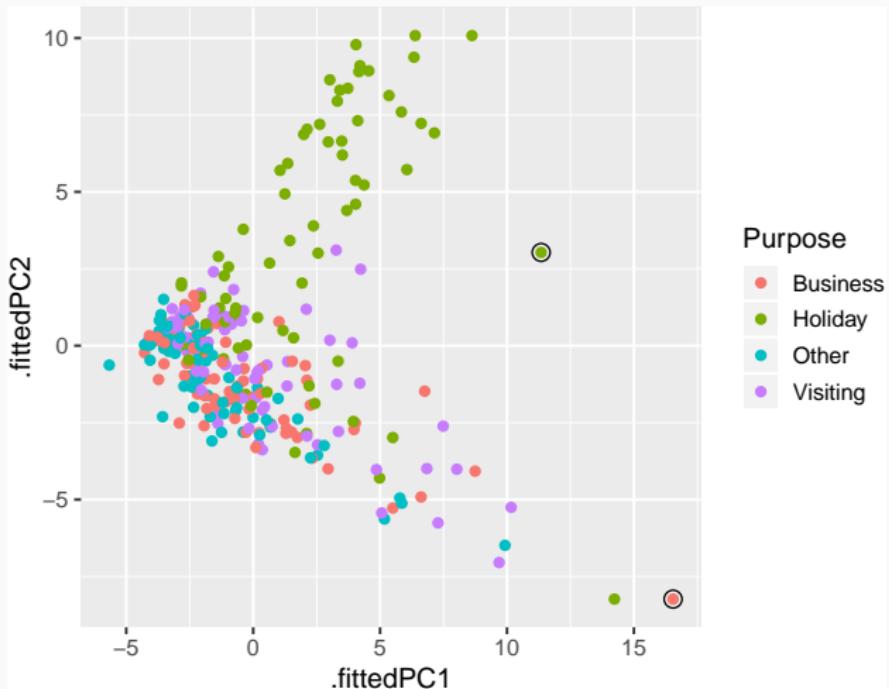
Principal components  
based on all features  
from the feasts  
package



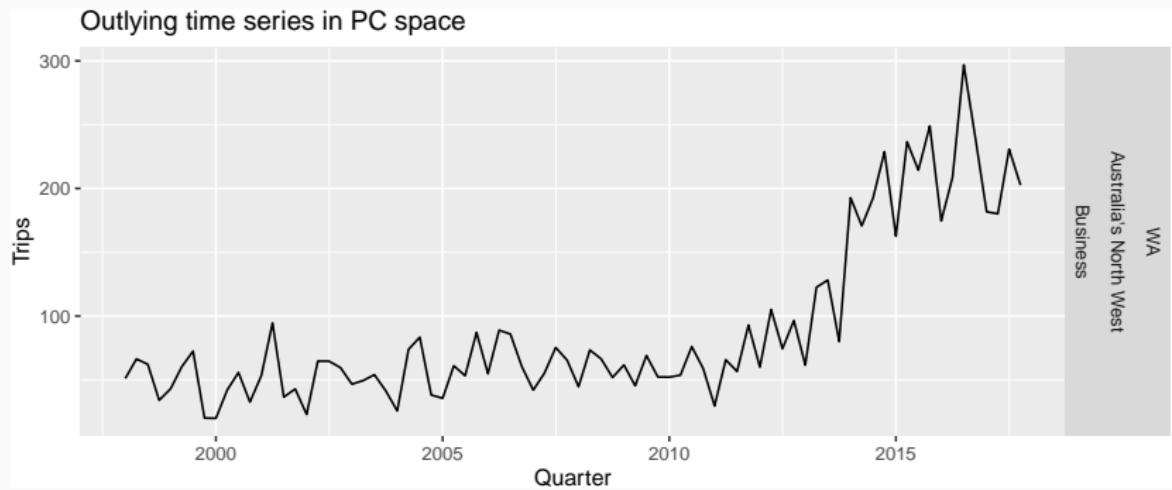
# Feature extraction and statistics

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

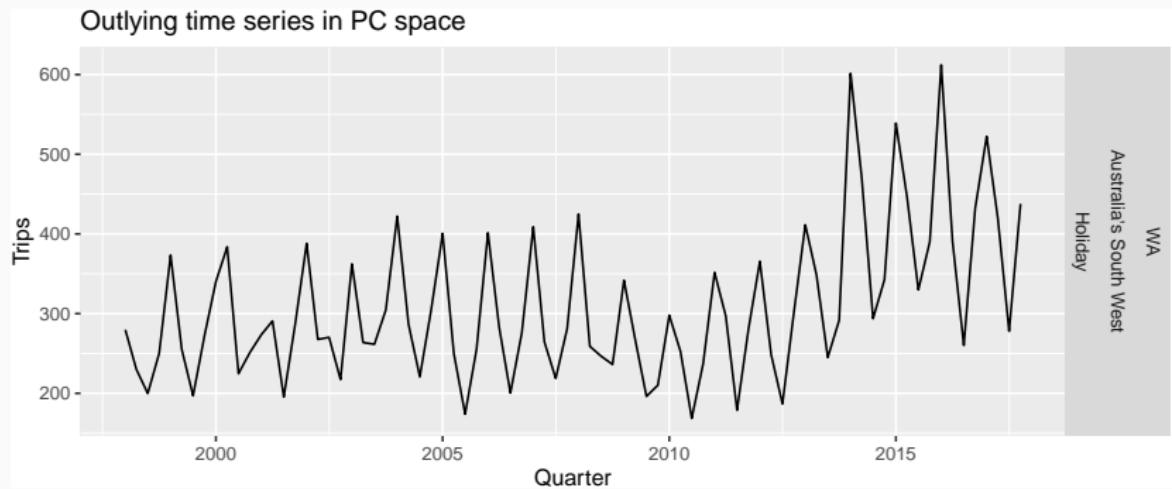
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics



# Feature extraction and statistics



# Outline

1 Visualization

2 R packages

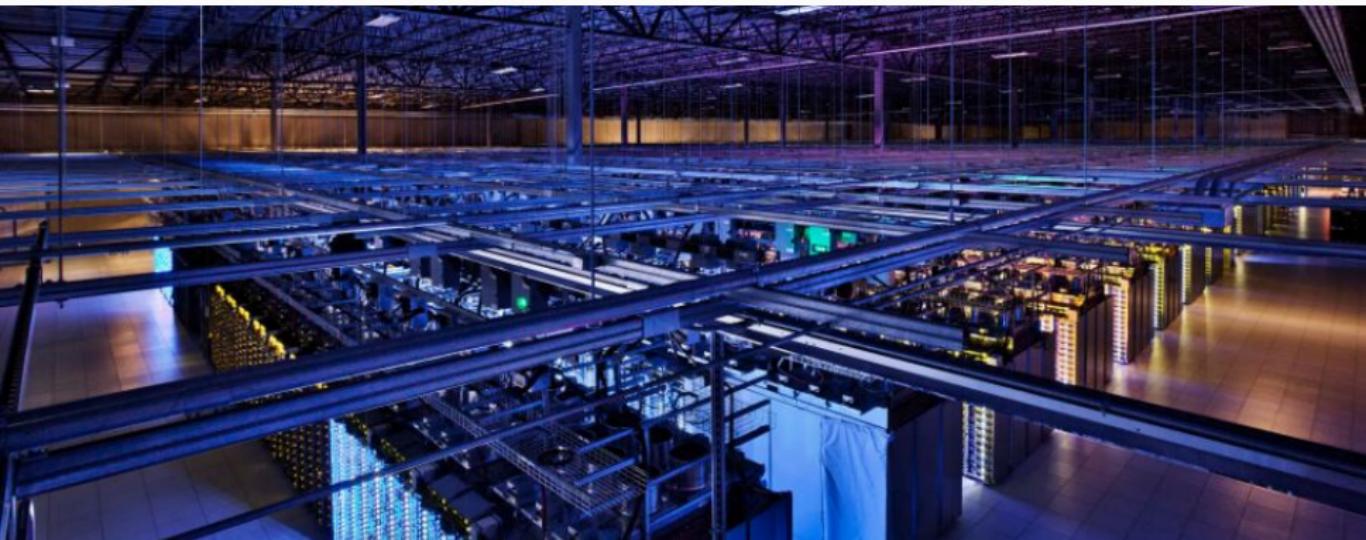
3 Anomaly detection

4 Forecast model selection

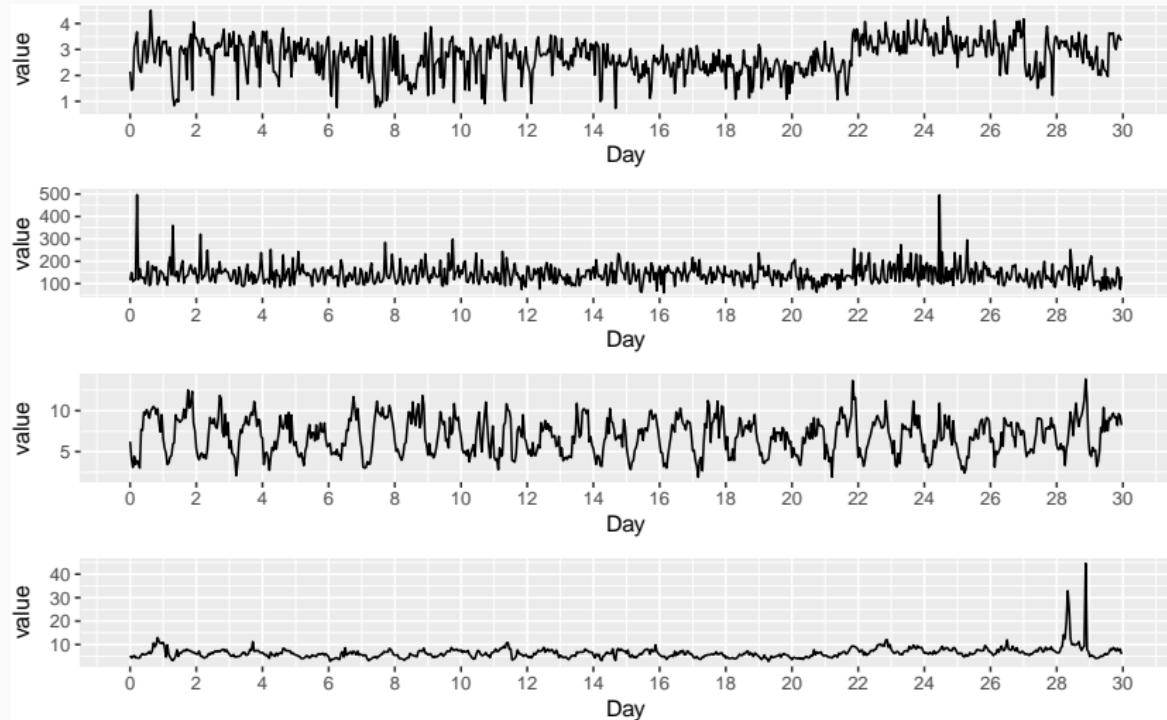
5 Forecast model averaging

# 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.



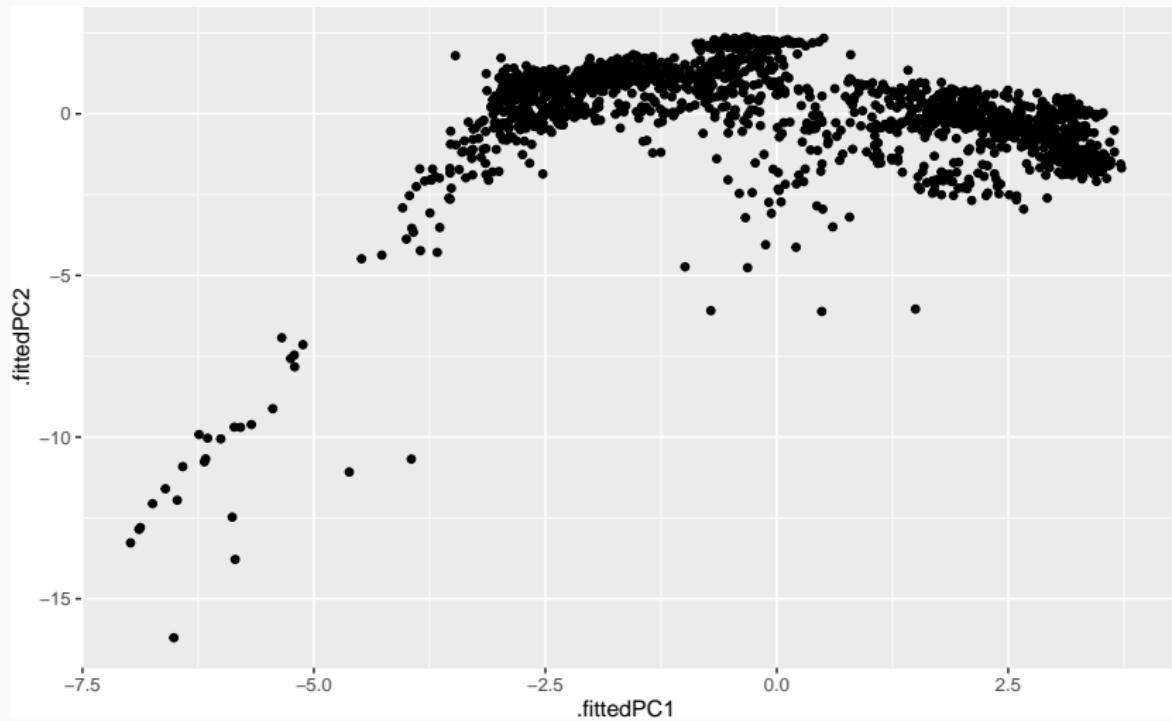
# Yahoo server metrics



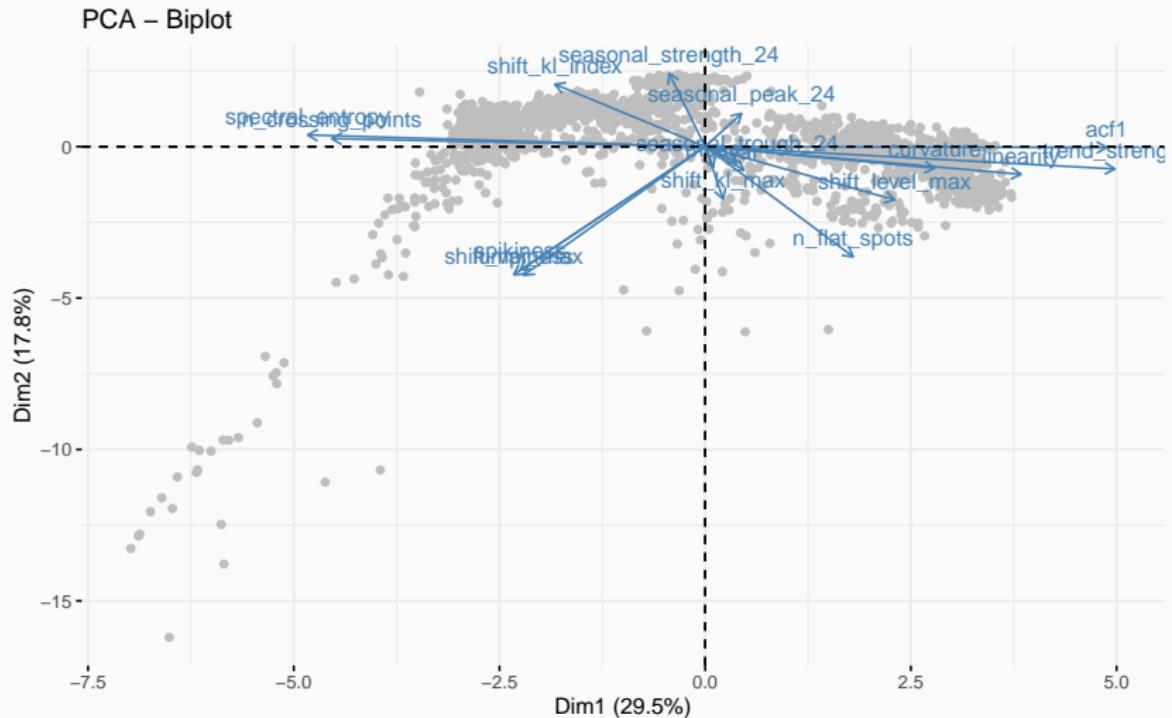
# Yahoo server metrics

- **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

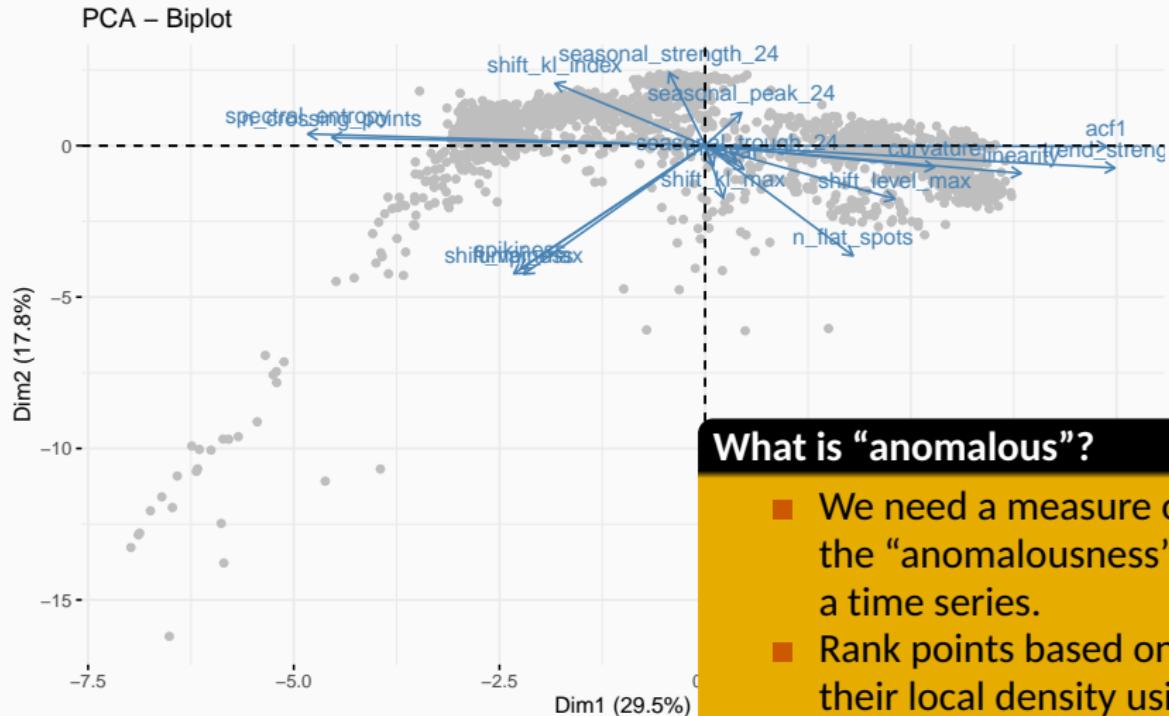
# Feature space



# Feature space



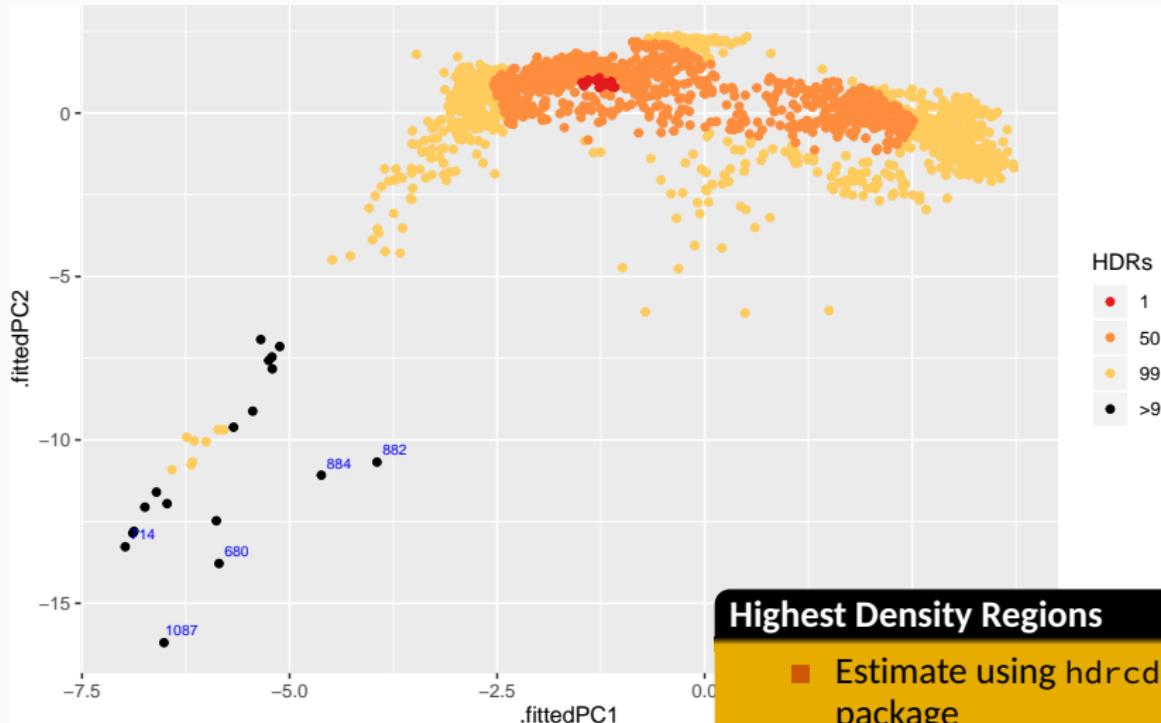
# Feature space



## 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 Visualization

2 R packages

3 Anomaly detection

4 Forecast model selection

5 Forecast model averaging

# Forecast model selection

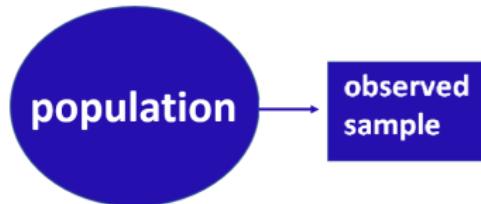
## Features used to select a forecasting model

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- first ACF value of remainder series
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive method
- unit root test statistics
- first ACF value of residual series of linear trend model
- ACF and PACF based features
  - calculated on both the raw and differenced series

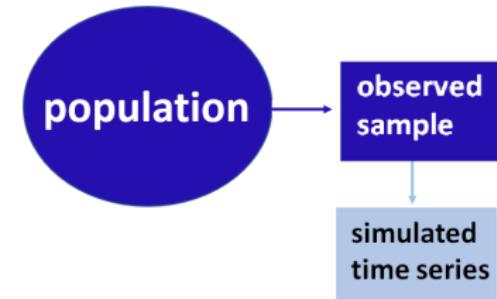
# FFORMS: Feature-based FORcast Model Selection

population

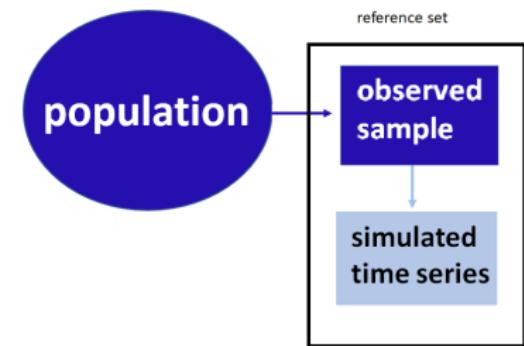
# FFORMS: Feature-based FORecast Model Selection



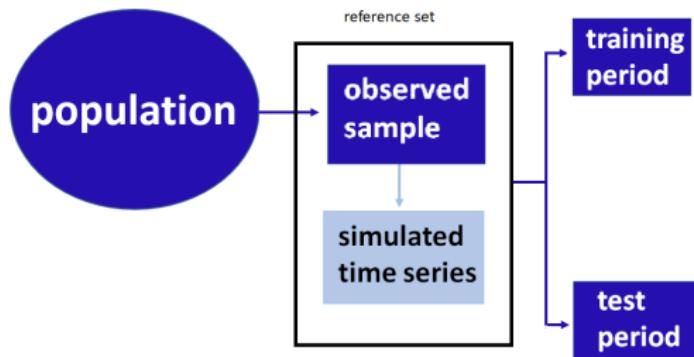
# FFORMS: Feature-based FORcast Model Selection



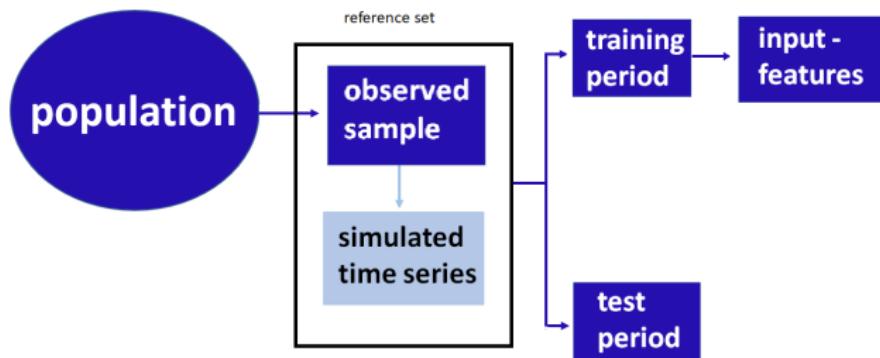
# FFORMS: Feature-based FOrecast Model Selection



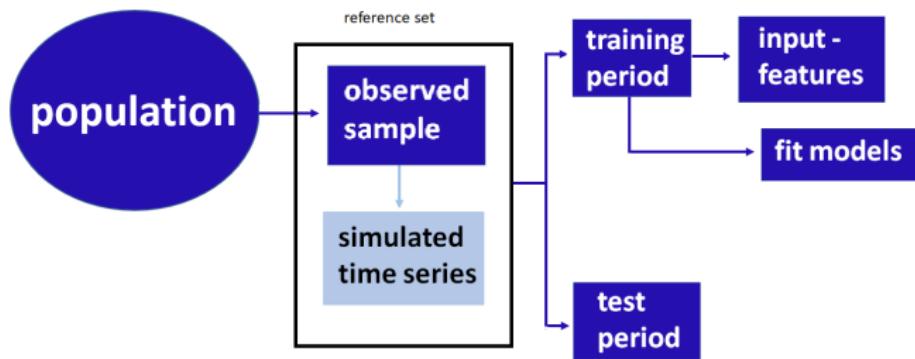
# FFORMS: Feature-based FORcast Model Selection



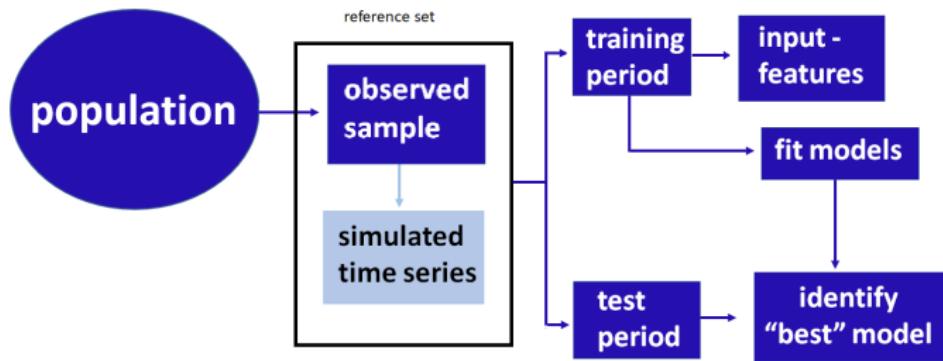
# FFORMS: Feature-based FORcast Model Selection



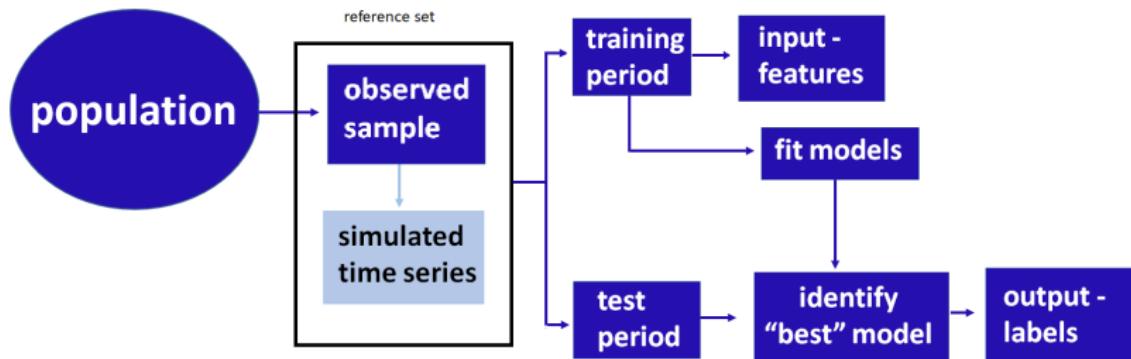
# FFORMS: Feature-based FORecast Model Selection



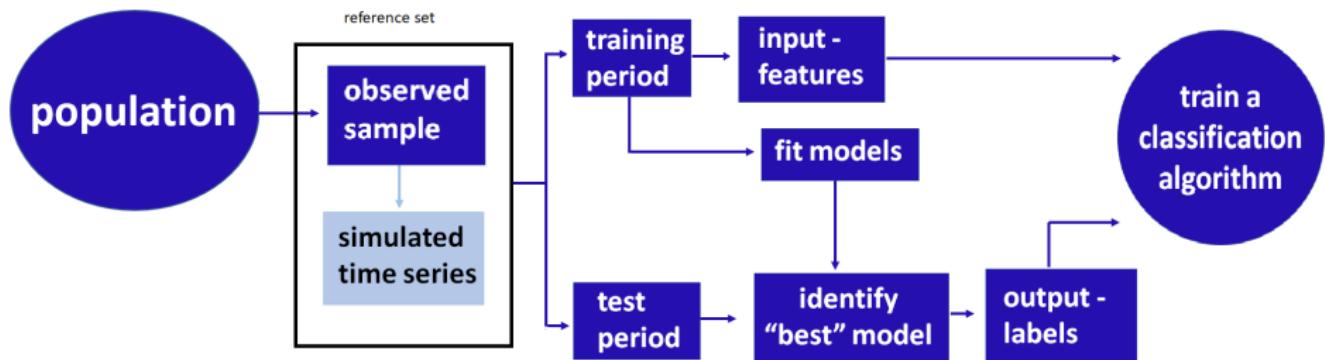
# FFORMS: Feature-based FORcast Model Selection



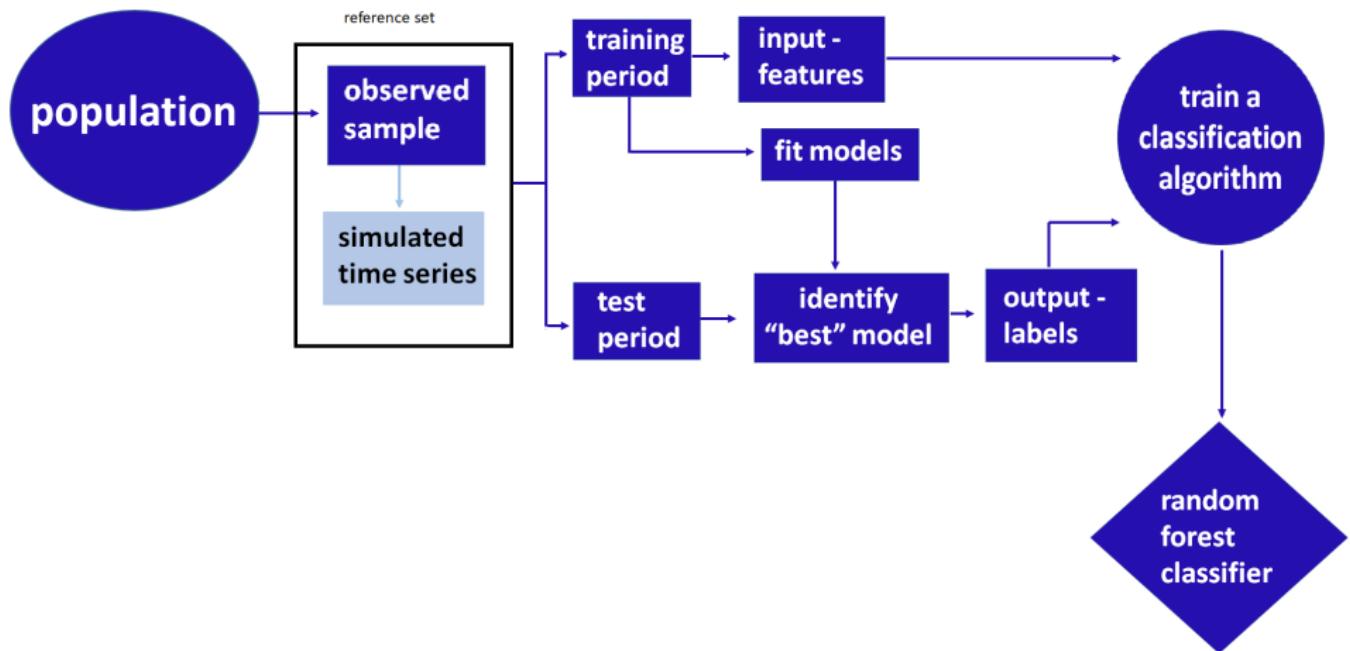
# FFORMS: Feature-based FORecast Model Selection



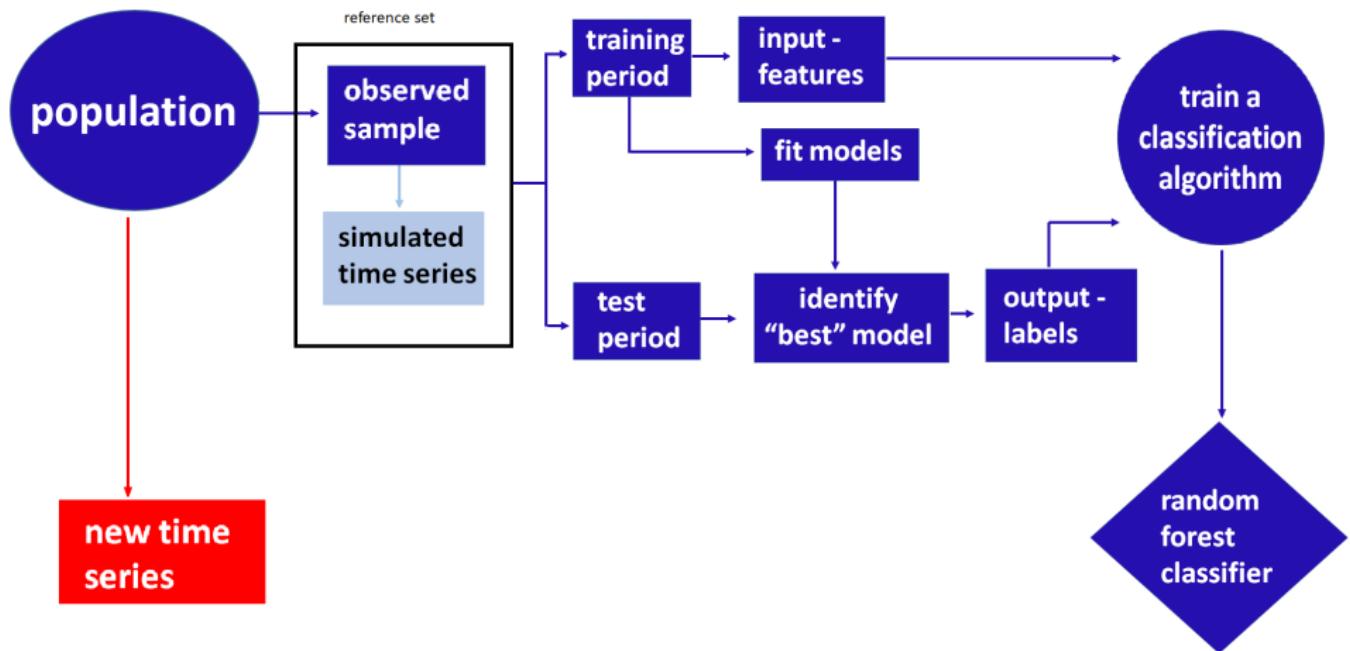
# FFORMS: Feature-based FORecast Model Selection



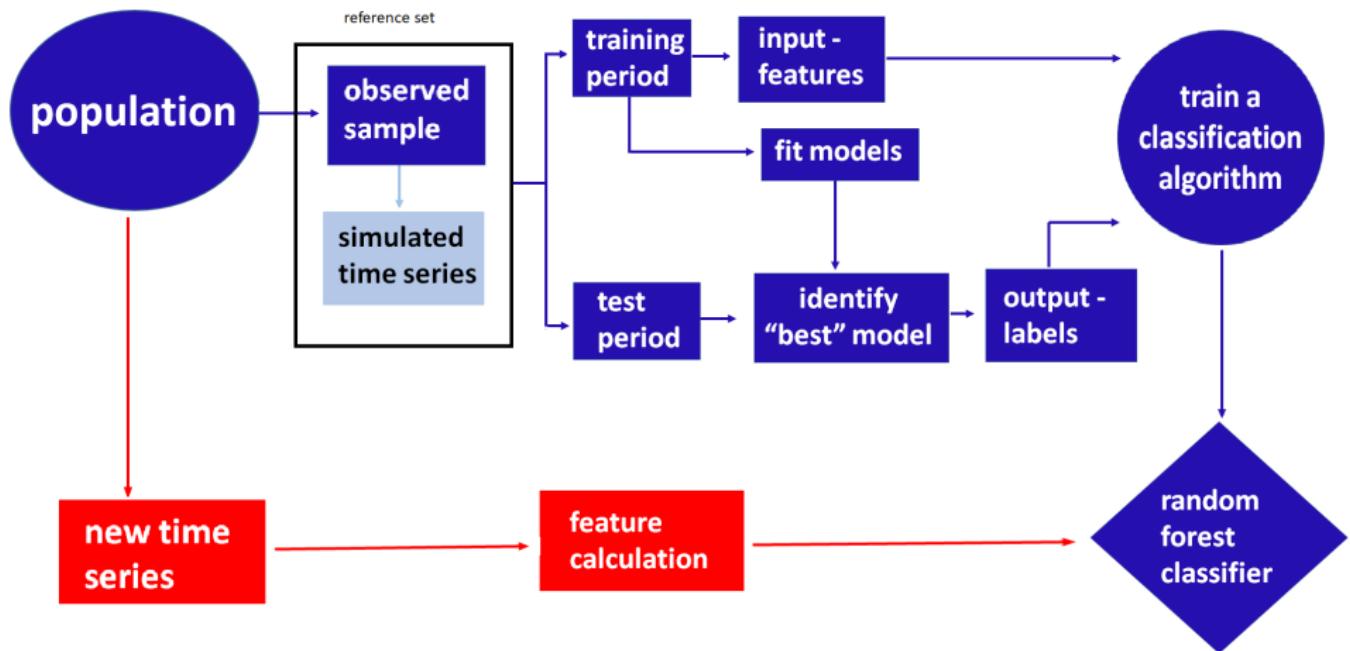
# FFORMS: Feature-based FORecast Model Selection



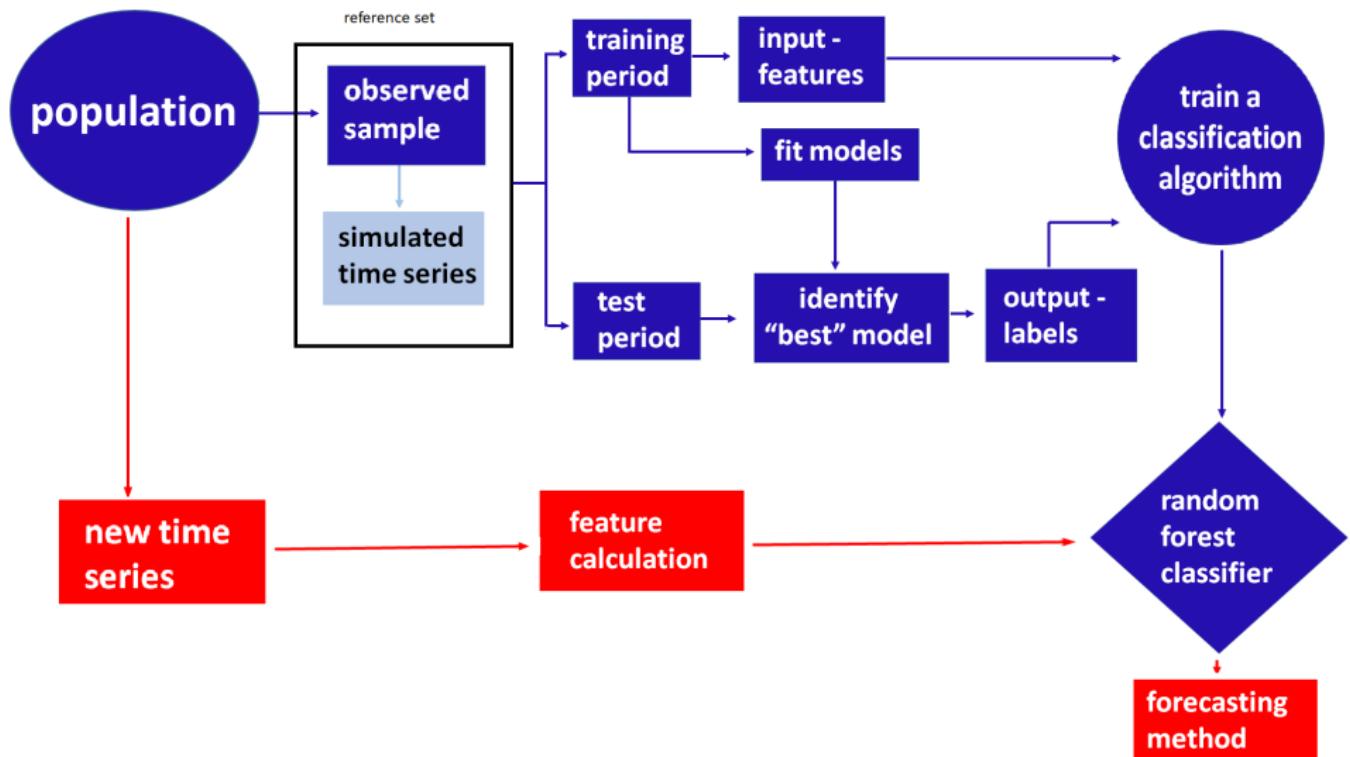
# FFORMS: Feature-based FORecast Model Selection



# FFORMS: Feature-based FORcast Model Selection



# FFORMS: Feature-based FORecast Model Selection



# Application to M competition data

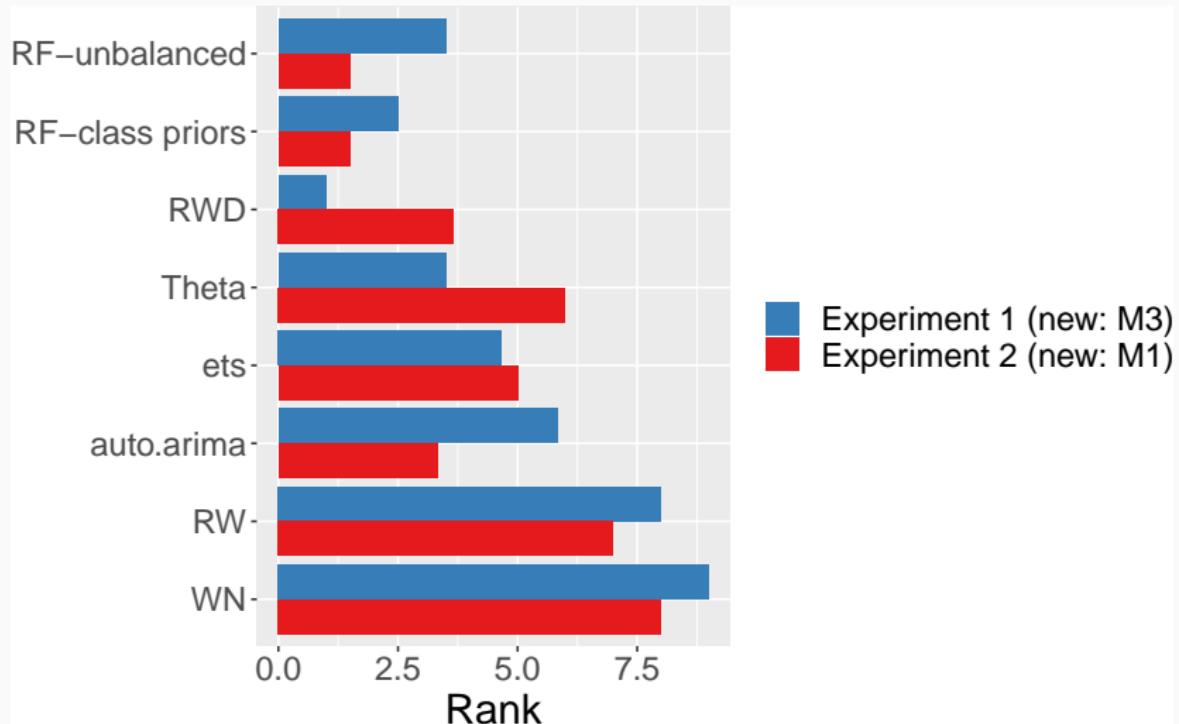
## Experiment 1

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

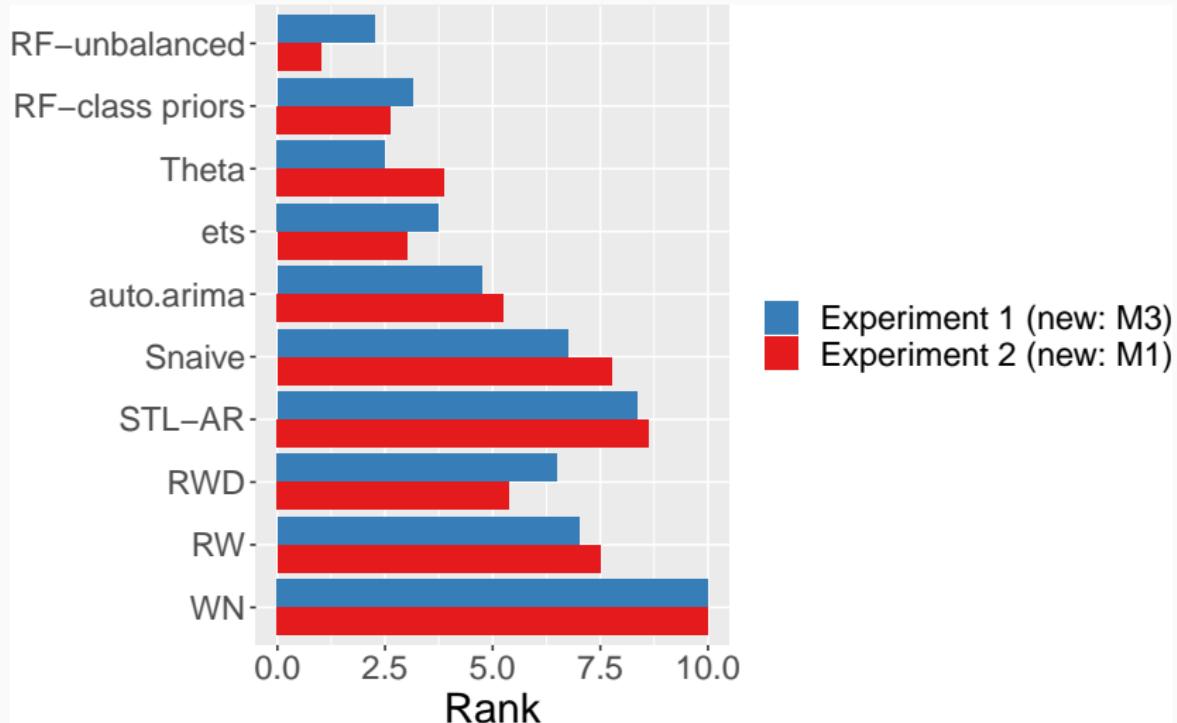
## Experiment 2

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

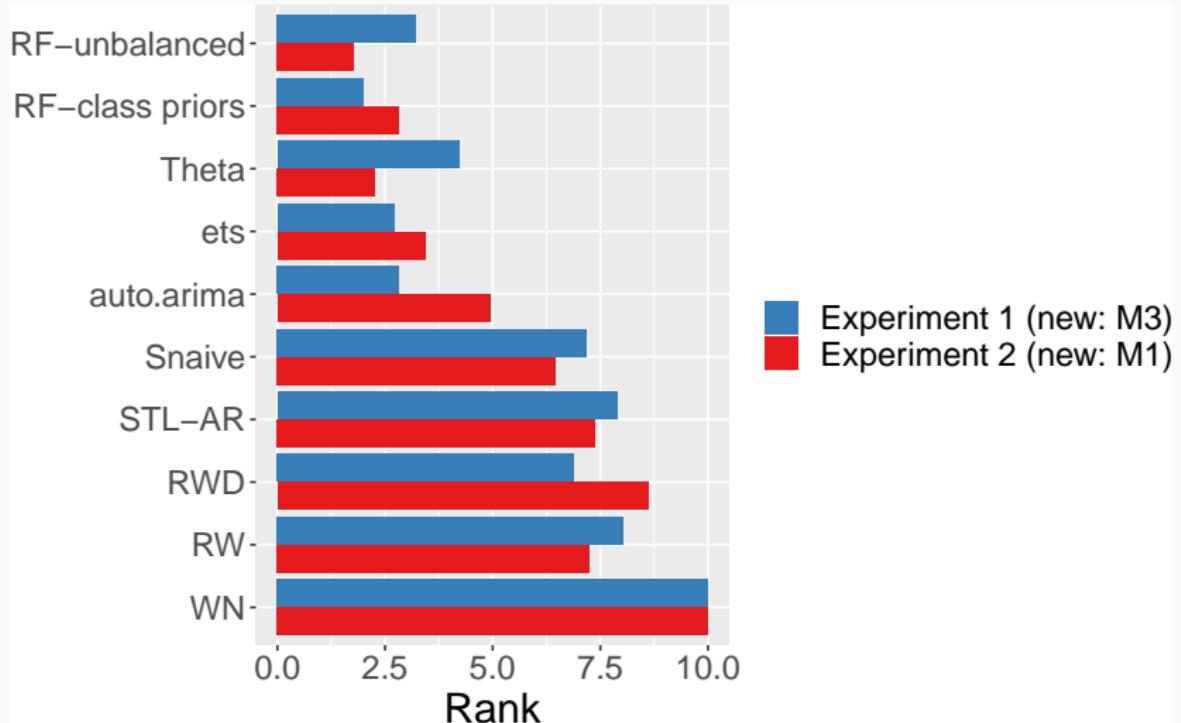
# Results: Yearly



# Results: Quarterly



# Results: Monthly



# Outline

1 Visualization

2 R packages

3 Anomaly detection

4 Forecast model selection

5 Forecast model averaging

## FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use gradient boosted trees rather than a random forest.
- The optimization criterion is forecast accuracy not classification accuracy.
- The probability of each model being best is used to construct a model weight.
- A combination forecast is produced using these weights.
- **Came second in the M4 forecasting competition**

# 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 STLM-AR
- 9 NNAR

# R Packages

- **seer**: FFORMS — selecting forecasting model using features.

`github.com/thiyangt/seer`

- **M4metalearning**: FFORMA – forecast combinations using features to choose weights.

`github.com/robjhyndman/M4metalearning`

# Acknowledgments



Dilini Talagala



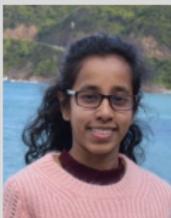
Earo Wang



Mitchell O'Hara-Wild



Kate Smith-Miles



Thiyanga Talagala



Yanfei Kang



George Athanasopoulos



Pablo Montero-Manso



Shanika Wickramasuriya