

# A feast of time series tools

Rob J Hyndman &  
Mitchell O'Hara-Wild

June 2019

# Outline

- 1 Overview
- 2 Tsibbles
- 3 Graphics
- 4 Decompositions
- 5 Features

# Outline

1 Overview

2 Tsibbles

3 Graphics

4 Decompositions

5 Features





## 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.

# Outline

1 Overview

2 Tsibbles

3 Graphics

4 Decompositions

5 Features

# tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
## # Key:      Country [263]
##   Year Country      GDP Imports Exports Population
##   <dbl> <fct>          <dbl>   <dbl>   <dbl>         <dbl>
## 1  1960 Afghanistan 5377777811.    7.02    4.13    8996351
## 2  1961 Afghanistan 5488888896.    8.10    4.45    9166764
## 3  1962 Afghanistan 5466666678.    9.35    4.88    9345868
## 4  1963 Afghanistan 7511111191.   16.9     9.17    9533954
## 5  1964 Afghanistan 8000000044.   18.1     8.89    9731361
## 6  1965 Afghanistan 10066666638.  21.4    11.3    9938414
## 7  1966 Afghanistan 13999999967.  18.6     8.57   10152331
## 8  1967 Afghanistan 16733333418.  14.2     6.77   10372630
## 9  1968 Afghanistan 13733333367.  15.2     8.90   10604346
## 10 1969 Afghanistan 14088888922.  15.0    10.1   10854428
## # ... with 15,140 more rows
```

# tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
```

```
## # Key:      Country [263]
```

```
##      Year Country      GDP Imports Exports Population
##      Index <fct>      <dbl>   <dbl>   <dbl>      <dbl>
##  1  1960 Afghanistan 5377777811.    7.02    4.13    8996351
##  2  1961 Afghanistan 5488888896.    8.10    4.45    9166764
##  3  1962 Afghanistan 5466666678.    9.35    4.88    9345868
##  4  1963 Afghanistan 7511111191.   16.9    9.17    9533954
##  5  1964 Afghanistan 8000000044.   18.1    8.89    9731361
##  6  1965 Afghanistan 10066666638.  21.4   11.3    9938414
##  7  1966 Afghanistan 13999999967.  18.6    8.57   10152331
##  8  1967 Afghanistan 16733333418.  14.2    6.77   10372630
##  9  1968 Afghanistan 13733333367.  15.2    8.90   10604346
## 10  1969 Afghanistan 14088888922.  15.0   10.1   10854428
## # ... with 15,140 more rows
```



# tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
```

```
## # Key:          Country [263]
```

```
##      Year Country      GDP Imports Exports Population
##      Index  Key      <dbl>   <dbl>   <dbl>         <dbl>
##  1  1960 Afghanistan 5377777811.    7.02    4.13    8996351
##  2  1961 Afghanistan 5488888896.    8.10    4.45    9166764
##  3  1962 Afghanistan 5466666678.    9.35    4.88    9345868
##  4  1963 Afghanistan 7511111191.   16.9    9.17    9533954
##  5  1964 Afghanistan 8000000044.   18.1    8.89    9731361
##  6  1965 Afghanistan 10066666638.   21.4   11.3    9938414
##  7  1966 Afghanistan 13999999967.   18.6    8.57   10152331
##  8  1967 Afghanistan 16733333418.   14.2    6.77   10372630
##  9  1968 Afghanistan 13733333367.   15.2    8.90   10604346
## 10  1969 Afghanistan 14088888922.   15.0   10.1   10854428
## # ... with 15,140 more rows
```

# tsibble objects

```
global_economy
```

```
## # A tsibble: 15,150 x 6 [1Y]
```

```
## # Key:      Country [263]
```

```
##      Year Country      GDP Imports Exports Population
```

```
##      Index  Key      Measured variables
```

```
## 1  1960 Afghanistan 5377777811.    7.02    4.13    8996351
```

```
## 2  1961 Afghanistan 5488888896.    8.10    4.45    9166764
```

```
## 3  1962 Afghanistan 5466666678.    9.35    4.88    9345868
```

```
## 4  1963 Afghanistan 7511111191.   16.9    9.17    9533954
```

```
## 5  1964 Afghanistan 8000000044.   18.1    8.89    9731361
```

```
## 6  1965 Afghanistan 10066666638.   21.4   11.3    9938414
```

```
## 7  1966 Afghanistan 13999999967.   18.6    8.57   10152331
```

```
## 8  1967 Afghanistan 16733333418.   14.2    6.77   10372630
```

```
## 9  1968 Afghanistan 13733333367.   15.2    8.90   10604346
```

```
## 10 1969 Afghanistan 14088888922.   15.0   10.1   10854428
```

```
## # ... with 15,140 more rows
```

# 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
```

# Holidays by state

```
holidays <- tourism %>%  
  filter(Purpose=="Visiting") %>%  
  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.
```

# Outline

1 Overview

2 Tsibbles

3 Graphics

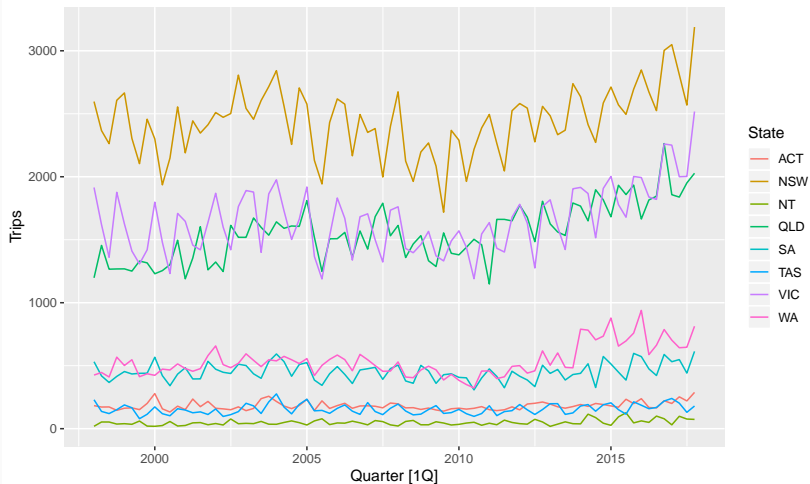
4 Decompositions

5 Features



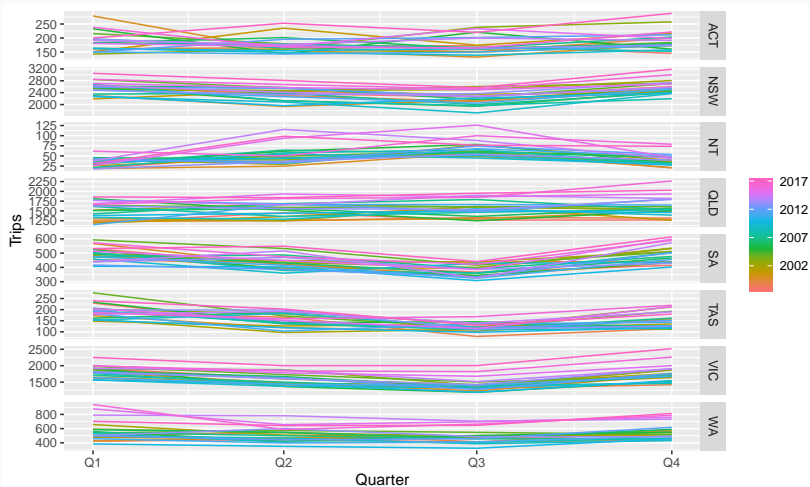
# Time plots

```
holidays %>% autoplot(Trips)
```



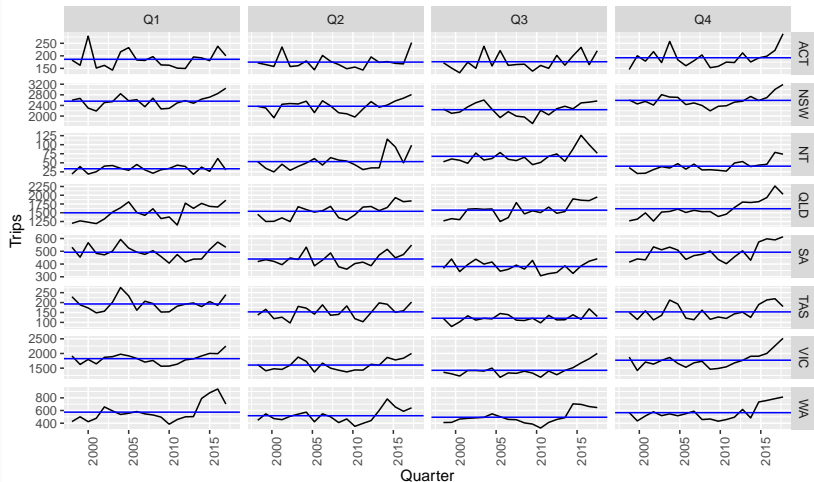
# Season plots

```
holidays %>% gg_season(Trips)
```



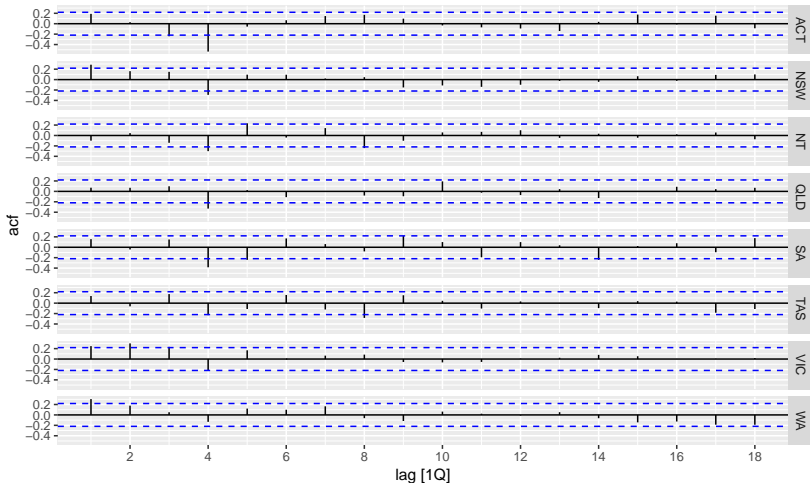
# Graphics

```
holidays %>% gg_subseries(Trips)
```



# Graphics

```
holidays %>% ACF(difference(Trips, 4)) %>% autoplot()
```



# Outline

- 1 Overview
- 2 Tsibbles
- 3 Graphics
- 4 Decompositions**
- 5 Features

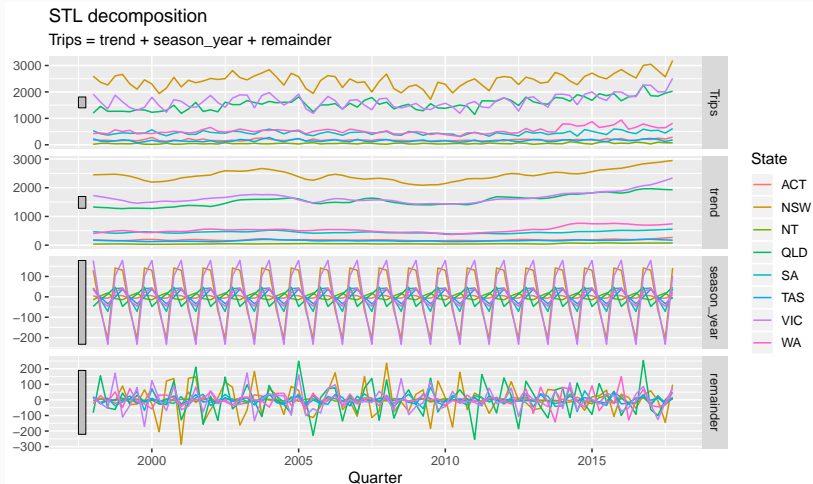
# Decompositions

The feasts package supports four common time series decomposition methods:

- Classical decomposition
- STL decomposition
- X11 decomposition
- X-13ARIMA-SEATS decomposition

# Decompositions

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



# Outline

- 1 Overview
- 2 Tsibbles
- 3 Graphics
- 4 Decompositions
- 5 Features



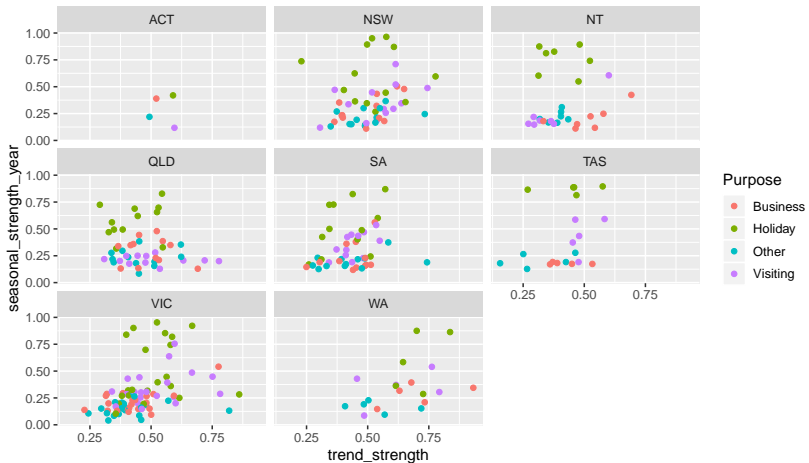
# Feature extraction and statistics

```
tourism %>% features(Trips, features_stl)
```

```
## # A tibble: 304 x 10
##   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 5 more variables:
## #   spike <dbl>, linearity <dbl>, curvature <dbl>,
## #   seasonal_peak_year <dbl>, seasonal_trough_year <dbl>
```

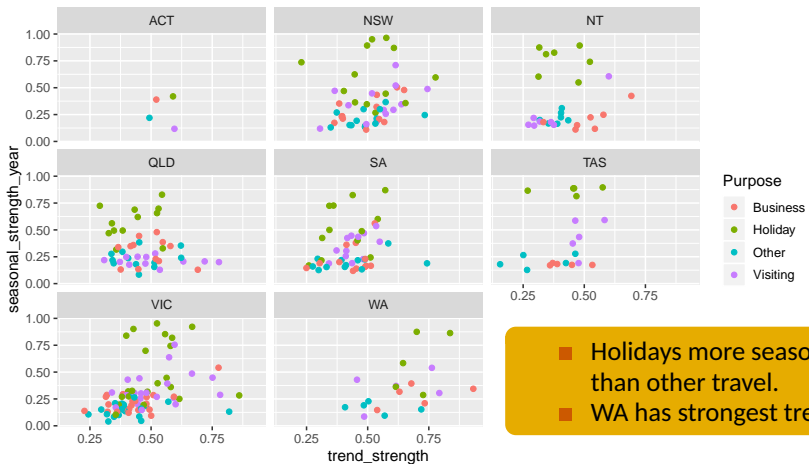
# Feature extraction and statistics

```
tourism %>% features(Trips, features_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, features_stl) %>%  
  ggplot(aes(x=trend_strength, y=seasonal_strength_year, col=Purpose)) +  
  geom_point() + facet_wrap(vars(State))
```



- Holidays more seasonal than other travel.
- WA has strongest trends.

# Feature extraction and statistics

Find the most seasonal time series:

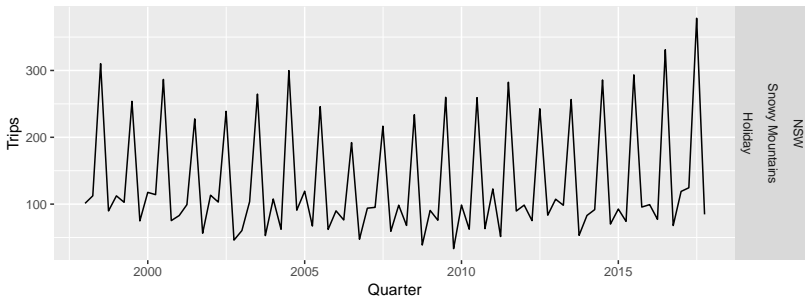
```
most_seasonal <- tourism %>%  
  features(Trips, features_stl) %>%  
  filter(seasonal_strength_year == max(seasonal_strength_year))
```

# Feature extraction and statistics

Find the most seasonal time series:

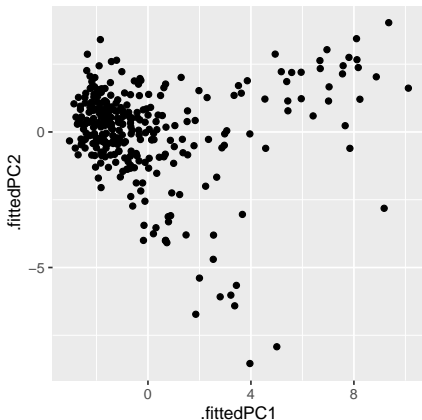
```
most_seasonal <- tourism %>%  
  features(Trips, features_stl) %>%  
  filter(seasonal_strength_year == max(seasonal_strength_year))
```

```
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

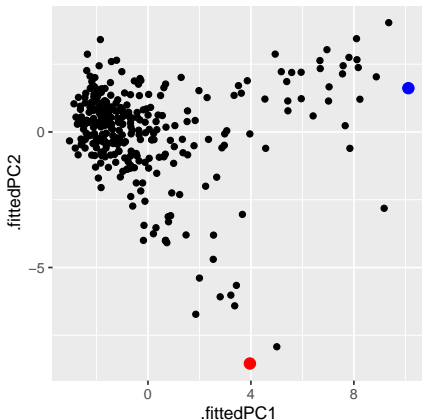
```
tourism_features <- tourism %>%  
  features(Trips, list(features_stl, features_acf, features_pacf))  
pcs <- tourism_features %>% select(-State, -Region, -Purpose) %>%  
  prcomp(scale=TRUE) %>% augment(tourism_features)  
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2)) + geom_point()
```



Principal components  
based on STL, ACF and  
PACF features

# Feature extraction and statistics

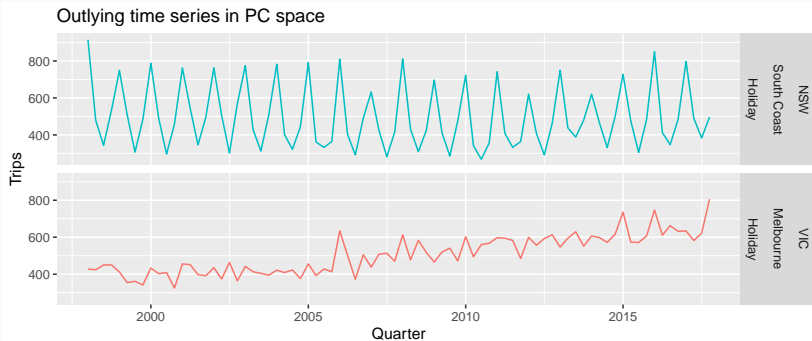
```
tourism_features <- tourism %>%  
  features(Trips, list(features_stl, features_acf, features_pacf))  
pcs <- tourism_features %>% select(-State, -Region, -Purpose) %>%  
  prcomp(scale=TRUE) %>% augment(tourism_features)  
pcs %>% ggplot(aes(x=.fittedPC1, y=.fittedPC2)) + geom_point()
```



Principal components  
based on STL, ACF and  
PACF features

# Feature extraction and statistics

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





# Acknowledgements



Mitchell O'Hara-Wild



Earo Wang

**[feasts.tidyverts.org](https://feasts.tidyverts.org)**  
**[robjhyndman.com](https://robjhyndman.com)**