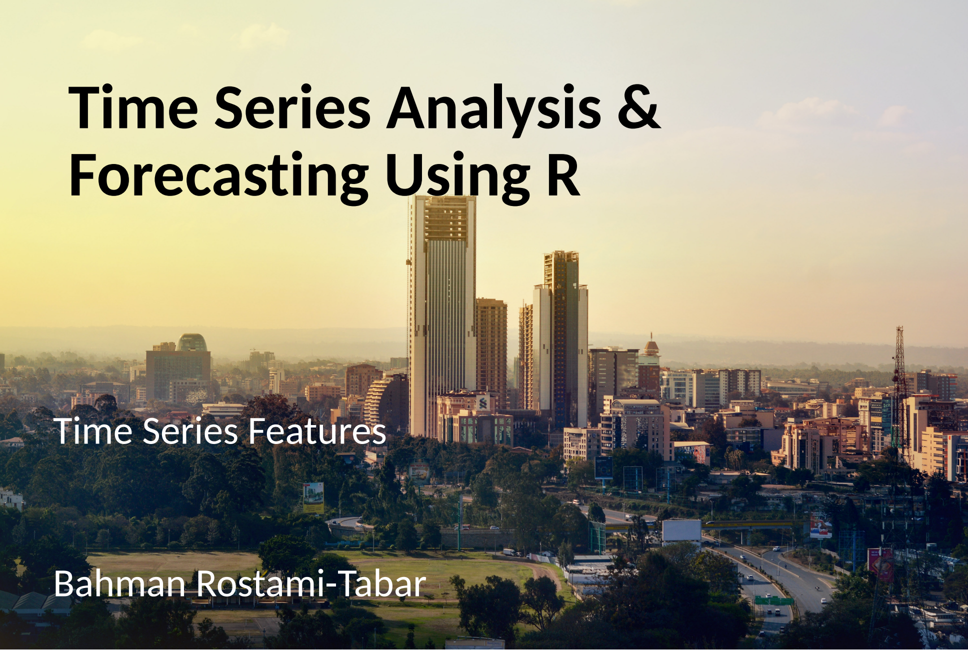


Time Series Analysis & Forecasting Using R

Time Series Features

Bahman Rostami-Tabar

Outline

1. Learning outcome
2. STL Features
3. Dimension reduction for features
4. Lab 3: time series features

Outline

Learning outcome

You should be able to:

1. Extract time series features (i.e. numerical summary) for a given time series
2. Visualise time series features for many time series
3. Use dimensional reduction techniques to understand data

Outline

Strength of seasonality and trend

STL decomposition

$$y_t = T_t + S_t + R_t$$

Seasonal strength

$$\max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right)$$

Trend strength

$$\max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)}\right)$$

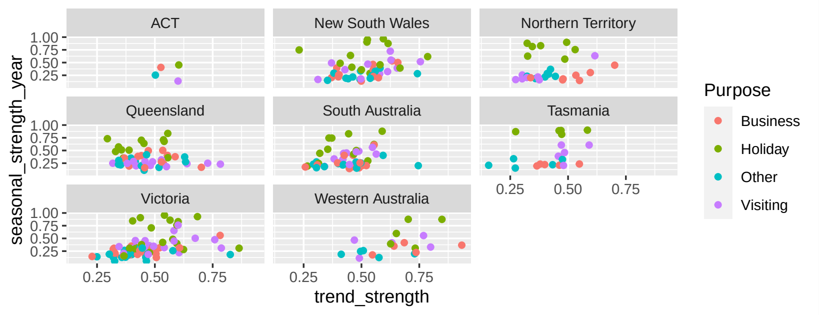
Feature extraction and statistics

```
tourism |> features(Trips, feat_stl)
```

```
## # A tibble: 304 x 12
##    Region    State Purpose trend~1 seaso~2 seaso~3
##    <chr>     <chr> <chr>     <dbl>   <dbl>   <dbl>
##  1 Adelaide  Sout~ Busine~   0.464   0.407       3
##  2 Adelaide  Sout~ Holiday   0.554   0.619       1
##  3 Adelaide  Sout~ Other     0.746   0.202       2
##  4 Adelaide  Sout~ Visiti~   0.435   0.452       1
##  5 Adelaide~ Sout~ Busine~   0.464   0.179       3
##  6 Adelaide~ Sout~ Holiday   0.528   0.296       2
##  7 Adelaide~ Sout~ Other     0.593   0.404       2
##  8 Adelaide~ Sout~ Visiti~   0.488   0.254       0
##  9 Alice Sp~ Nort~ Busine~   0.534   0.251       0
## 10 Alice Sp~ Nort~ Holiday   0.381   0.832       3
## # ... with 294 more rows, 6 more variables:
## #   seasonal_trough_year <dbl>, spikiness <dbl>,
## #   linearity <dbl>, curvature <dbl>,
## #   stl_e_acf1 <dbl>, stl_e_acf10 <dbl>, and
```

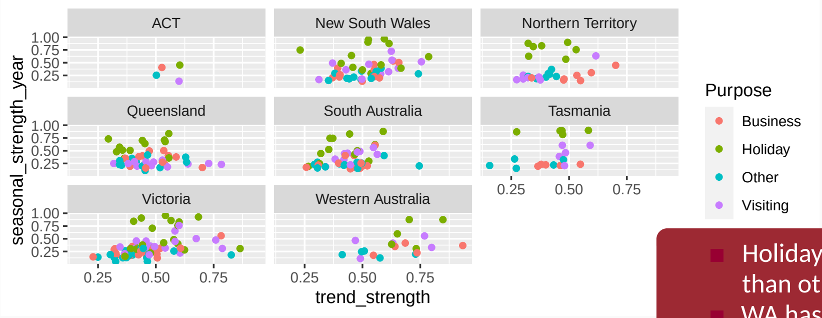
Feature extraction and statistics

```
tourism |>
  features(Trips, feat_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, feat_stl) |>
  ggplot(aes(x = trend_strength, y = seasonal_strength_year, col = Purpose)) +
  geom_point() + facet_wrap(vars(State))
```



- Holidays more
than other tra
- WA has strong

Feature extraction and statistics

Find the most seasonal time series:

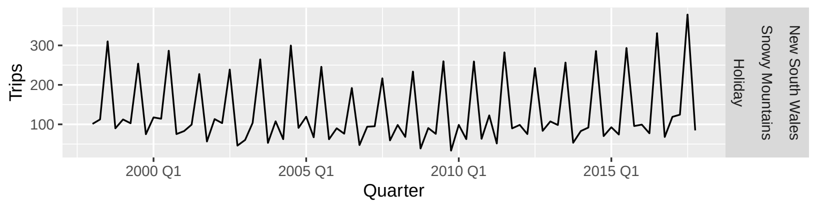
```
most_seasonal <- tourism |>
  features(Trips, feat_stl) |>
  filter(seasonal_strength_year == max(seasonal_strength_year))
```

Feature extraction and statistics

Find the most seasonal time series:

```
most_seasonal <- tourism |>
  features(Trips, feat_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

Find the most trended time series:

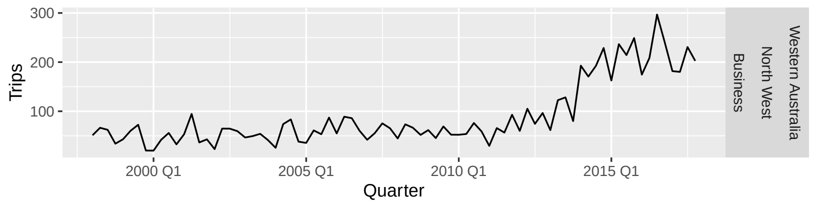
```
most_trended <- tourism |>
  features(Trips, feat_stl) |>
  filter(trend_strength == max(trend_strength))
```

Feature extraction and statistics

Find the most trended time series:

```
most_trended <- tourism |>
  features(Trips, feat_stl) |>
  filter(trend_strength == max(trend_strength))
```

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



Outline

Feature extraction and statistics

```r
tourism_features <- tourism |>
  features(Trips, feature_set(pkgs = "feasts"))
```

All features from the feasts package

```
## # A tibble: 304 x 51
##    Region    State Purpose trend~1 seaso~2 seaso~3
##    <chr>     <chr> <chr>     <dbl>   <dbl>   <dbl>
##  1 Adelaide  Sout~ Busine~   0.464   0.407       3
##  2 Adelaide  Sout~ Holiday   0.554   0.619       1
##  3 Adelaide  Sout~ Other     0.746   0.202       2
##  4 Adelaide  Sout~ Visiti~   0.435   0.452       1
##  5 Adelaide~ Sout~ Busine~   0.464   0.179       3
##  6 Adelaide~ Sout~ Holiday   0.528   0.296       2
##  7 Adelaide~ Sout~ Other     0.593   0.404       2
##  8 Adelaide~ Sout~ Visiti~   0.488   0.254       0
##  9 Alice Sp~ Nort~ Busine~   0.534   0.251       0
## 10 Alice Sp~ Nort~ Holiday   0.381   0.832       3
## # ... with 294 more rows, 45 more variables:
## #   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>,
```

Feature extraction and statistics

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

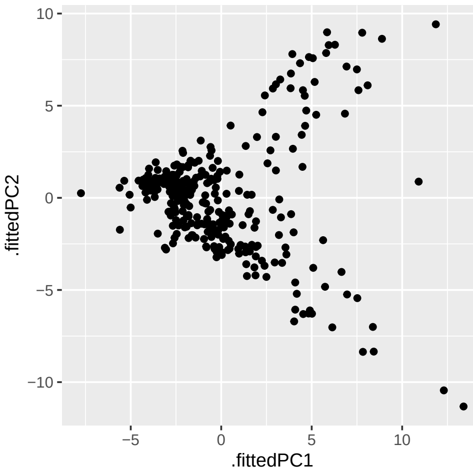
Principal components b
all features from the fe
package

```
## # A tibble: 304 x 100
##    .rownames Region  State Purpose trend~1 seaso~2
##    <chr>     <chr>   <chr> <chr>     <dbl>   <dbl>
##  1 1         Adelai~ Sout~ Busine~   0.464   0.407
##  2 2         Adelai~ Sout~ Holiday   0.554   0.619
##  3 3         Adelai~ Sout~ Other     0.746   0.202
##  4 4         Adelai~ Sout~ Visiti~   0.435   0.452
##  5 5         Adelai~ Sout~ Busine~   0.464   0.179
##  6 6         Adelai~ Sout~ Holiday   0.528   0.296
##  7 7         Adelai~ Sout~ Other     0.593   0.404
##  8 8         Adelai~ Sout~ Visiti~   0.488   0.254
##  9 9         Alice ~ Nort~ Busine~   0.534   0.251
## 10 10        Alice ~ Nort~ Holiday   0.381   0.832
## # ... with 294 more rows, 94 more variables:
## #   seasonal_peak_year <dbl>,
## #   seasonal_trough_year <dbl>, spikiness <dbl>,
```

Feature extraction and statistics

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

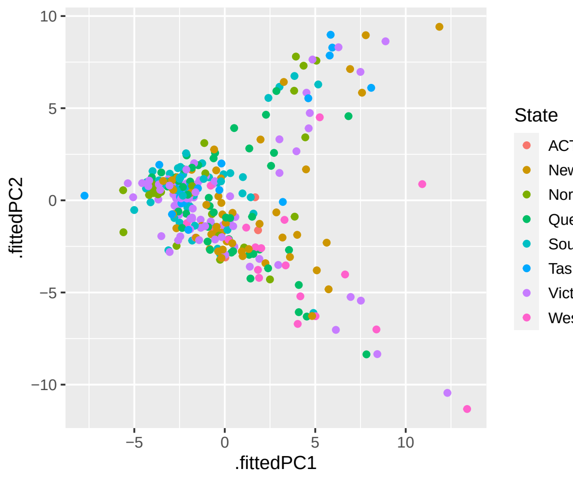
Principal components based on all features from the feasts package



Feature extraction and statistics

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

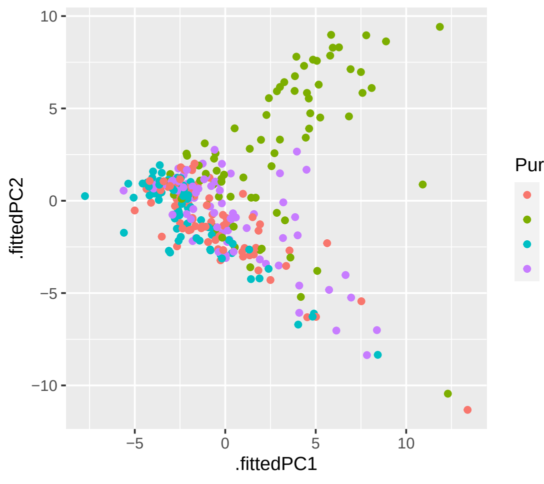
Principal components based on all features from the feasts package



Feature extraction and statistics

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

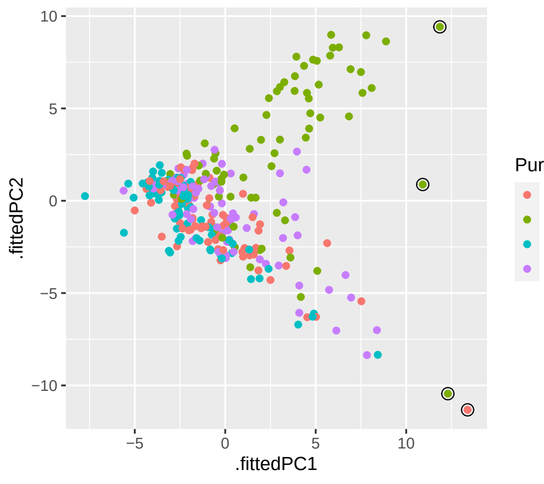
Principal components based on all features from the feasts package



Feature extraction and statistics

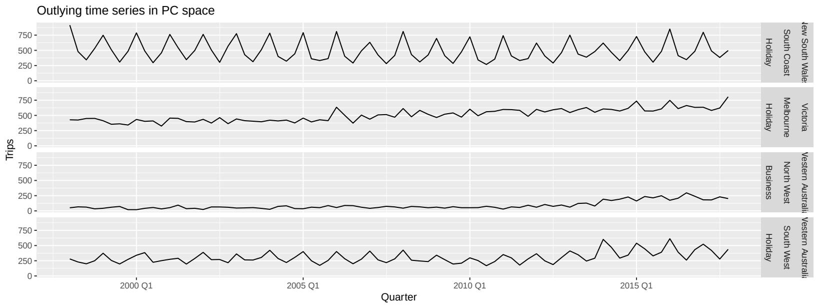
```
pcs |> ggplot(aes(x=.fittedPC1, y=.fittedPC2, col=Purpose)) +
  geom_point() + theme(aspect.ratio=1)
```

Principal components based on all features from the feasts package



Feature extraction and statistics

```r
outliers |>
  left_join(tourism, by = c("State", "Region", "Purpose")) |>
  mutate(Series = glue("{State}", "{Region}", "{Purpose}", .sep = "\n\n")) |>
  ggplot(aes(x = Quarter, y = Trips)) +
  geom_line() + facet_grid(Series ~ .) +
  labs(title = "Outlying time series in PC space")
```



Outline

Lab 3: time series features

- for the daily admissions time series with keys (with gender, injury), extract the strength of trend and seasonality
  - Do you see any useful insight?
- Use `GGally::ggpairs()` to look at the relationships between the STL-based features. You might wish to change `seasonal_peak_year` and `seasonal_trough_year` to factors.
- Which is the peak quarter for holidays in each state?
- Use a feature-based approach to look for outlying series in `PBS`.