

# Exploratory time series analysis using R



# Outline

- 1 STL Features
- 2 Dimension reduction for features
- 3 Lab Session 4

# Outline

- 1 STL Features
- 2 Dimension reduction for features
- 3 Lab Session 4

# 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)
```

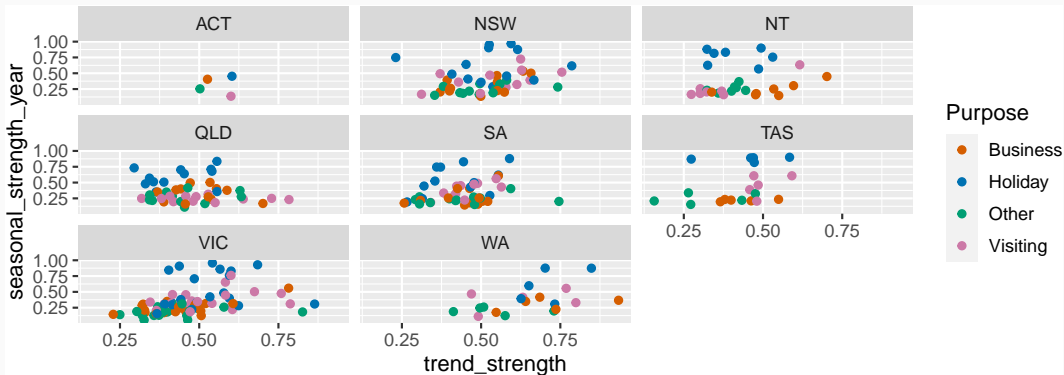
```
## # A tibble: 304 x 12
##   Region      State Purpose trend~1 season~2 season~3 season~4 spiki~5 linea~6 curva~7
##   <chr>      <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Adelaide SA      Busine~  0.464    0.407      3        1 1.58e+2  -5.31    71.6
## 2 Adelaide SA      Holiday 0.554    0.619      1        2 9.17e+0  49.0     78.7
## 3 Adelaide SA      Other   0.746    0.202      2        1 2.10e+0  95.1     43.4
## 4 Adelaide SA      Visiti~ 0.435    0.452      1        3 5.61e+1  34.6     71.4
## 5 Adelaide Hi~ SA      Busine~  0.464    0.179      3        0 1.03e-1  0.968   -3.22
## 6 Adelaide Hi~ SA      Holiday 0.528    0.296      2        1 1.77e-1  10.5     24.0
## 7 Adelaide Hi~ SA      Other   0.593    0.404      2        2 4.44e-4  4.28     3.19
## 8 Adelaide Hi~ SA      Visiti~ 0.488    0.254      0        3 6.50e+0  34.2    -0.529
## 9 Alice Sprin~ NT      Busine~  0.534    0.251      0        1 1.69e-1  23.8     19.5
## 10 Alice Sprin~ NT      Holiday 0.381    0.832      3        1 7.39e-1 -19.6     10.5
## # ... with 294 more rows, 2 more variables: stl_e_acf1 <dbl>, stl_e_acf10 <dbl>, and
## # abbreviated variable names 1: trend_strength, 2: seasonal_strength_year,
## # 3: seasonal_peak_year, 4: seasonal_trough_year, 5: spikiness, 6: linearity
```

# 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))
```

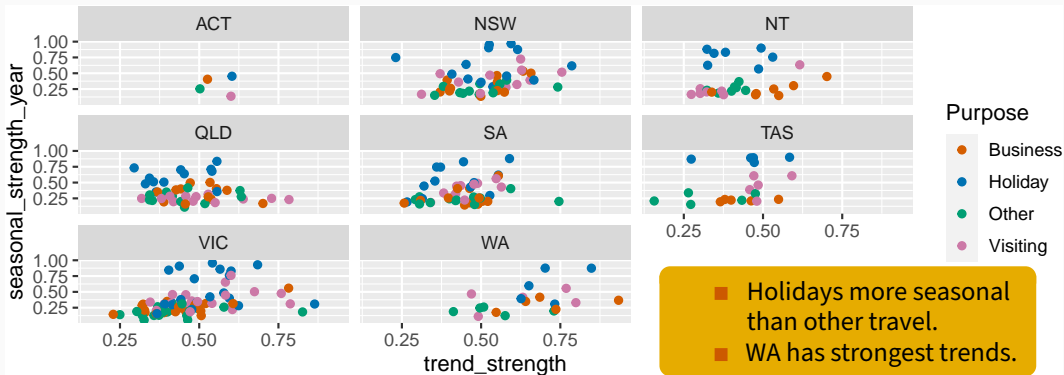


# Feature extraction and statistics

tourism ▷

```
features(Trips, feat_stl) ▷
```

```
ggplot(aes(x = trend_strength, y = seasonal_strength_year, col = Purpose)) +  
geom_point() + facet_wrap(vars(State))
```



# 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))
```

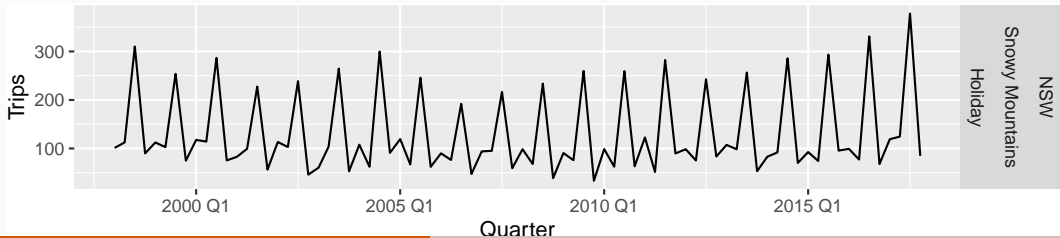


# Feature extraction and statistics

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))
```



# Outline

- 1 STL Features
- 2 Dimension reduction for features
- 3 Lab Session 4

# Feature extraction and statistics

```
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 season~2 season~3 season~4 spiki~5 linea~6 curva~7  
##   <chr>      <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 Adelaide SA      Busine~ 0.464    0.407      3        1 1.58e+2 -5.31    71.6  
## 2 Adelaide SA      Holiday 0.554    0.619      1        2 9.17e+0 49.0     78.7  
## 3 Adelaide SA      Other   0.746    0.202      2        1 2.10e+0 95.1     43.4  
## 4 Adelaide SA      Visiti~ 0.435    0.452      1        3 5.61e+1 34.6     71.4  
## 5 Adelaide Hi~ SA      Busine~ 0.464    0.179      3        0 1.03e-1 0.968    -3.22  
## 6 Adelaide Hi~ SA      Holiday 0.528    0.296      2        1 1.77e-1 10.5     24.0  
## 7 Adelaide Hi~ SA      Other   0.593    0.404      2        2 4.44e-4 4.28     3.19  
## 8 Adelaide Hi~ SA      Visiti~ 0.488    0.254      0        3 6.50e+0 34.2     -0.529  
## 9 Alice Sprin~ NT      Busine~ 0.534    0.251      0        1 1.69e-1 23.8     19.5  
## 10 Alice Sprin~ NT      Holiday 0.381    0.832      3        1 7.39e-1 -19.6     10.5  
## # ... with 294 more rows, 41 more variables: stl_e_acf1 <dbl>, stl_e_acf10 <dbl>,  
## #   acf1 <dbl>, acf10 <dbl>, diff1_acf1 <dbl>, diff1_acf10 <dbl>, diff2_acf1 <dbl>,  
## #   diff2_acf10 <dbl>, season_acf1 <dbl>, pacf5 <dbl>, diff1_pacf5 <dbl>,  
## #   ...
```

# Feature extraction and statistics

```
colnames(tourism_features)
```

```
## [1] "Region"          "State"           "Purpose"
## [4] "trend_strength"  "seasonal_strength_year" "seasonal_peak_year"
## [7] "seasonal_trough_year" "spikiness"       "linearity"
## [10] "curvature"       "stl_e_acf1"      "stl_e_acf10"
## [13] "acf1"            "acf10"           "diff1_acf1"
## [16] "diff1_acf10"     "diff2_acf1"      "diff2_acf10"
## [19] "season_acf1"     "pacf5"           "diff1_pacf5"
## [22] "diff2_pacf5"     "season_pacf"     "zero_run_mean"
## [25] "nonzero_squared_cv" "zero_start_prop" "zero_end_prop"
## [28] "lambda_guerrero" "kpss_stat"       "kpss_pvalue"
## [31] "pp_stat"         "pp_pvalue"       "ndiffs"
## [34] "nsdiffs"         "bp_stat"         "bp_pvalue"
## [37] "lb_stat"         "lb_pvalue"       "var_tiled_var"
## [40] "var_tiled_mean"  "shift_level_max" "shift_level_index"
## [43] "shift_var_max"   "shift_var_index" "shift_kl_max"
## [46] "shift_kl_index"  "spectral_entropy" "n_crossing_points"
## [49] "longest_flat_spot" "coef_hurst"      "stat_arch_lm"
```

# Feature extraction and statistics

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

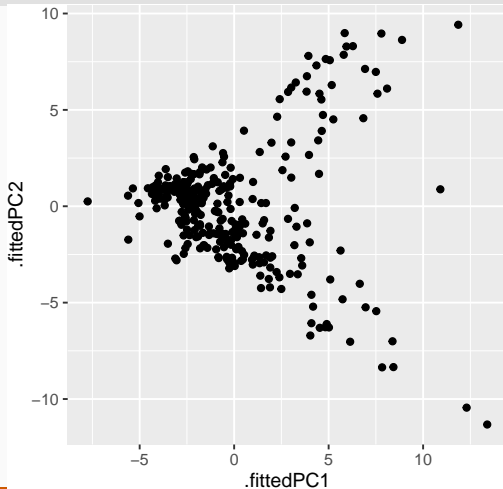
Principal components based  
on all features from the feasts  
package

```
## # A tibble: 304 x 100  
##   .rownames Region      State Purpose trend~1 season~2 season~3 season~4 spiki~5 lineac~6  
##   <chr>      <chr>      <chr> <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>  
## 1 1      Adelaide SA      Busine~ 0.464      0.407      3          1 1.58e+2 -5.31  
## 2 2      Adelaide SA      Holiday 0.554      0.619      1          2 9.17e+0 49.0  
## 3 3      Adelaide SA      Other   0.746      0.202      2          1 2.10e+0 95.1  
## 4 4      Adelaide SA      Visiti~ 0.435      0.452      1          3 5.61e+1 34.6  
## 5 5      Adelaide ~ SA      Busine~ 0.464      0.179      3          0 1.03e-1 0.968  
## 6 6      Adelaide ~ SA      Holiday 0.528      0.296      2          1 1.77e-1 10.5  
## 7 7      Adelaide ~ SA      Other   0.593      0.404      2          2 4.44e-4 4.28  
## 8 8      Adelaide ~ SA      Visiti~ 0.488      0.254      0          3 6.50e+0 34.2  
## 9 9      Alice Spr~ NT      Busine~ 0.534      0.251      0          1 1.69e-1 23.8  
## 10 10     Alice Spr~ NT      Holiday 0.381      0.832      3          1 7.39e-1 -19.6  
## # ... with 294 more rows, 90 more variables: curvature <dbl>, stl_e_acf1 <dbl>,
```

# Feature extraction and statistics

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

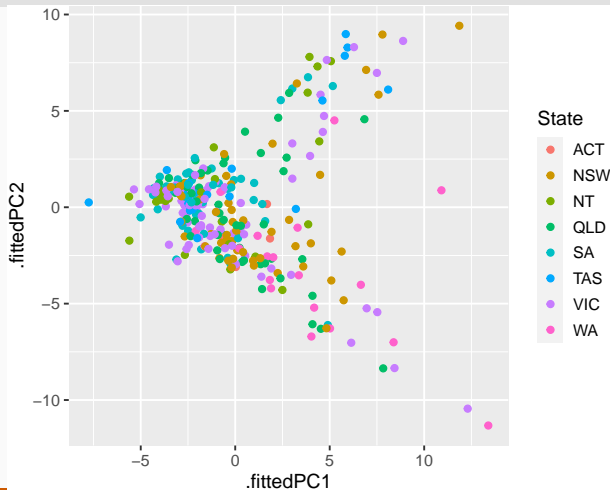
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

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

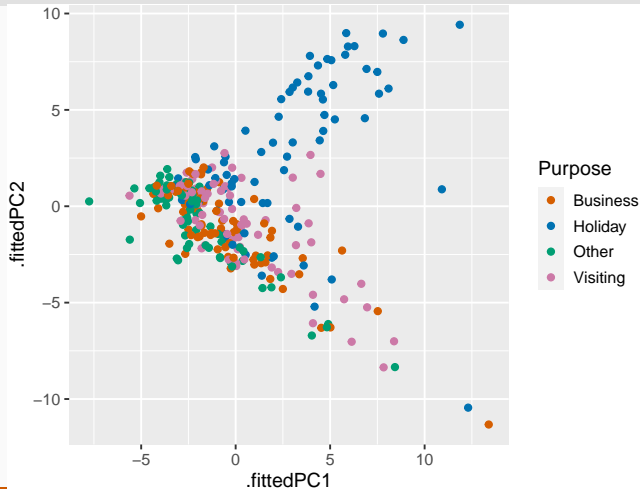
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

```
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

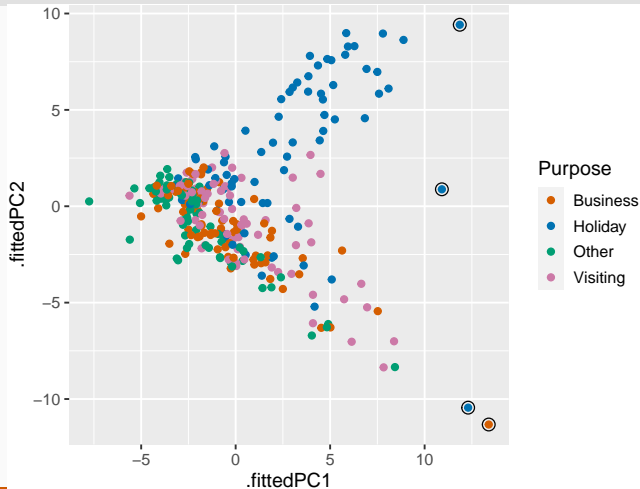




# Feature extraction and statistics

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

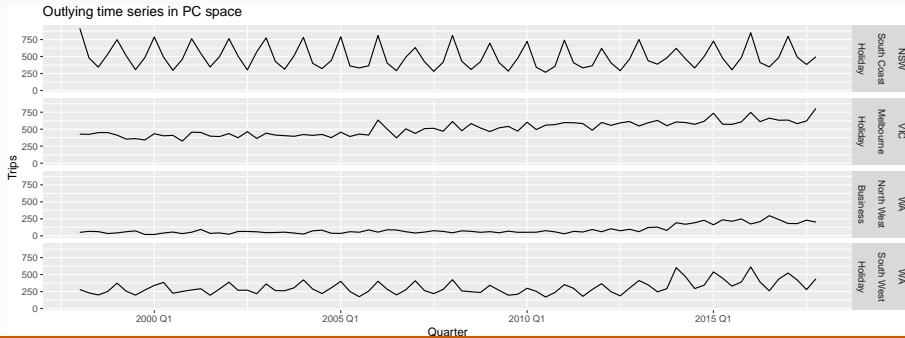
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

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

- 1 STL Features
- 2 Dimension reduction for features
- 3 Lab Session 4

## Lab Session 4

- Use a feature-based approach to look for outlying series in PBS.
- What is unusual about the series you identify as outliers?