

# 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

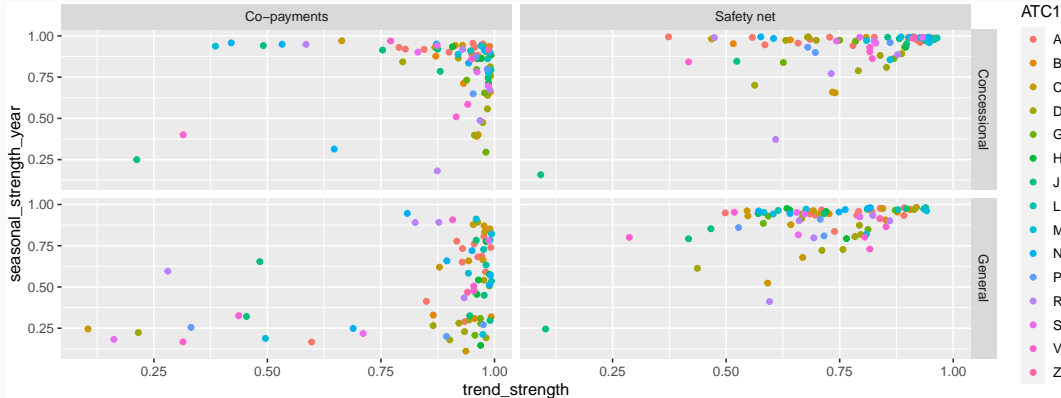
PBS ▷

```
features(Scripts, feat_stl)
```

```
## # A tibble: 336 x 13
##   Conces~1 Type  ATC1  ATC2  trend~2  season~3  season~4  season~5  spiki~6  linea~7  curva~8
##   <chr>    <chr> <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Concess~ Co-p~ A    A01    0.845    0.918      9        7 2.03e 7 -2.22e4 -5165.
## 2 Concess~ Co-p~ A    A02    0.970    0.943     11        6 6.44e13 2.09e6 27300.
## 3 Concess~ Co-p~ A    A03    0.956    0.931      9        7 1.61e 9 -1.07e5 -56979.
## 4 Concess~ Co-p~ A    A04    0.790    0.930      9        7 3.85e 9 -5.95e4 5758.
## 5 Concess~ Co-p~ A    A05    0.950    0.888     11        6 3.18e 3 2.14e3 -695.
## 6 Concess~ Co-p~ A    A06    0.952    0.955      9        6 7.18e 8 9.46e4 -33858.
## 7 Concess~ Co-p~ A    A07    0.804    0.920      9        6 6.64e 8 -2.86e4 -34172.
## 8 Concess~ Co-p~ A    A09    0.900    0.901     11        6 2.29e 4 4.69e3 -1702.
## 9 Concess~ Co-p~ A    A10    0.975    0.952     11        6 3.40e12 1.06e6 21547.
## 10 Concess~ Co-p~ A    A11    0.963    0.881     11        7 1.35e 8 -5.66e4 -39186.
## # ... with 326 more rows, 2 more variables: stl_e_acf1 <dbl>, stl_e_acf10 <dbl>, and 5
## # abbreviated variable names 1: Concession 2: trend strength
```

# Feature extraction and statistics

```
PBS > features(Scripts, feat_stl) >  
ggplot(aes(x = trend_strength, y = seasonal_strength_year, col=ATC1)) +  
geom_point() + facet_grid(Concession ~ Type)
```



# Feature extraction and statistics

Find the most seasonal time series:

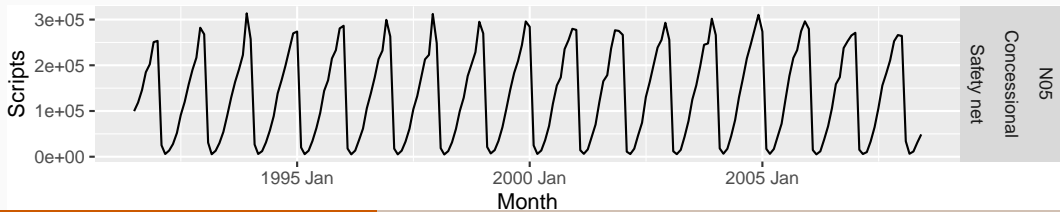
```
most_seasonal <- PBS ▷  
  features(Scripts, feat_stl) ▷  
  arrange(desc(seasonal_strength_year)) ▷  
  head(1)
```

# Feature extraction and statistics

Find the most seasonal time series:

```
most_seasonal <- PBS ▷  
  features(Scripts, feat_stl) ▷  
  arrange(desc(seasonal_strength_year)) ▷  
  head(1)
```

```
PBS ▷  
  semi_join(most_seasonal, by = c("ATC1", "ATC2", "Concession", "Type")) ▷  
  ggplot(aes(x = Month, y = Scripts)) +  
  geom_line() + facet_grid(vars(ATC2, Concession, Type))
```





# Feature extraction and statistics

Find the most trended time series:

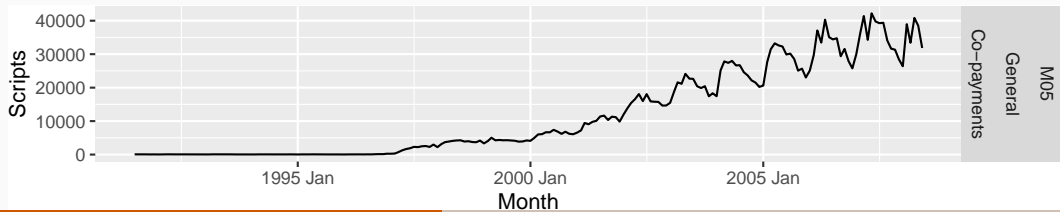
```
most_trended <- PBS ▷  
  features(Scripts, feat_stl) ▷  
  arrange(desc(trend_strength)) ▷  
  head(1)
```

# Feature extraction and statistics

Find the most trended time series:

```
most_trended <- PBS ▷  
  features(Scripts, feat_stl) ▷  
  arrange(desc(trend_strength)) ▷  
  head(1)
```

```
PBS ▷  
  semi_join(most_trended, by = c("ATC1", "ATC2", "Concession", "Type")) ▷  
  ggplot(aes(x = Month, y = Scripts)) +  
  geom_line() + facet_grid(vars(ATC2, Concession, Type))
```



# Outline

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

# Feature extraction and statistics

```
PBS_features <- PBS ▷  
  features(Scripts, feature_set(pkgs = "feasts")) ▷  
  select(-`...26`) ▷  
  na.omit()
```

All features from the feasts package

```
## # A tibble: 333 x 52  
##   Conces~1 Type ATC1 ATC2 trend~2 season~3 season~4 season~5 spiki~6 linea~7 curva~8  
##   <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Concess~ Co-p~ A A01 0.845 0.918 9 7 2.03e 7 -2.22e4 -5165.  
## 2 Concess~ Co-p~ A A02 0.970 0.943 11 6 6.44e13 2.09e6 27300.  
## 3 Concess~ Co-p~ A A03 0.956 0.931 9 7 1.61e 9 -1.07e5 -56979.  
## 4 Concess~ Co-p~ A A04 0.790 0.930 9 7 3.85e 9 -5.95e4 5758.  
## 5 Concess~ Co-p~ A A05 0.950 0.888 11 6 3.18e 3 2.14e3 -695.  
## 6 Concess~ Co-p~ A A06 0.952 0.955 9 6 7.18e 8 9.46e4 -33858.  
## 7 Concess~ Co-p~ A A07 0.804 0.920 9 6 6.64e 8 -2.86e4 -34172.  
## 8 Concess~ Co-p~ A A09 0.900 0.901 11 6 2.29e 4 4.69e3 -1702.  
## 9 Concess~ Co-p~ A A10 0.975 0.952 11 6 3.40e12 1.06e6 21547.  
## 10 Concess~ Co-p~ A A11 0.963 0.881 11 7 1.35e 8 -5.66e4 -39186. 10  
## # ... with 323 more rows, 41 more variables: stl_e_acf1 <dbl>, stl_e_acf10 <dbl>,
```

# Feature extraction and statistics

```
colnames(PBS_features)
```

```
## [1] "Concession"      "Type"            "ATC1"
## [4] "ATC2"            "trend_strength"  "seasonal_strength_year"
## [7] "seasonal_peak_year" "seasonal_trough_year" "spikiness"
## [10] "linearity"       "curvature"       "stl_e_acf1"
## [13] "stl_e_acf10"     "acf1"            "acf10"
## [16] "diff1_acf1"      "diff1_acf10"     "diff2_acf1"
## [19] "diff2_acf10"     "season_acf1"     "pacf5"
## [22] "diff1_pacf5"     "diff2_pacf5"     "season_pacf"
## [25] "zero_run_mean"   "nonzero_squared_cv" "zero_start_prop"
## [28] "zero_end_prop"   "lambda_guerrero" "kpss_stat"
## [31] "kpss_pvalue"     "pp_stat"         "pp_pvalue"
## [34] "ndiffs"          "nsdiffs"         "bp_stat"
## [37] "bp_pvalue"       "lb_stat"         "lb_pvalue"
## [40] "var_tiled_var"   "var_tiled_mean"  "shift_level_max"
## [43] "shift_level_index" "shift_var_max"   "shift_var_index"
## [46] "shift_kl_max"    "shift_kl_index"  "spectral_entropy"
## [49] "n_crossing_points" "longest_flat_spot" "coef_hurst"
## [52] "stat_arch_lm"
```

# Feature extraction and statistics

```
pcs <- PBS_features ▷  
  select(-ATC1, -ATC2, -Type, -Concession) ▷  
  prcomp(scale = TRUE) ▷  
  broom::augment(PBS_features)
```

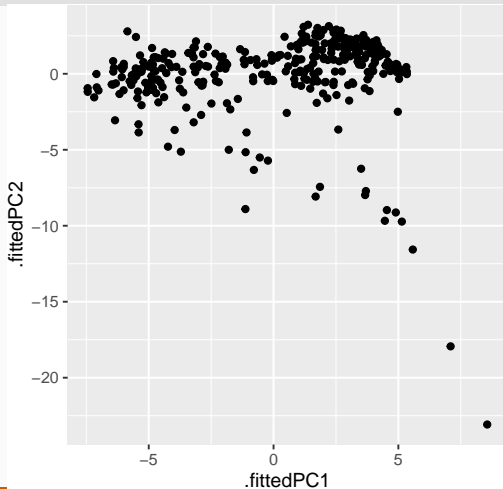
Principal components based  
on all features from the feasts  
package

```
## # A tibble: 333 x 101  
##   .rowna~1 Conce~2 Type  ATC1  ATC2  trend~3  season~4  season~5  season~6  spiki~7  line~8  
##   <chr>    <chr>    <chr> <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 1      Conces~ Co-p~ A    A01    0.845    0.918      9      7 2.03e 7 -2.22e4  
## 2 2      Conces~ Co-p~ A    A02    0.970    0.943     11      6 6.44e13 2.09e6  
## 3 3      Conces~ Co-p~ A    A03    0.956    0.931      9      7 1.61e 9 -1.07e5  
## 4 4      Conces~ Co-p~ A    A04    0.790    0.930      9      7 3.85e 9 -5.95e4  
## 5 5      Conces~ Co-p~ A    A05    0.950    0.888     11      6 3.18e 3 2.14e3  
## 6 6      Conces~ Co-p~ A    A06    0.952    0.955      9      6 7.18e 8 9.46e4  
## 7 7      Conces~ Co-p~ A    A07    0.804    0.920      9      6 6.64e 8 -2.86e4  
## 8 8      Conces~ Co-p~ A    A09    0.900    0.901     11      6 2.29e 4 4.69e3  
## 9 9      Conces~ Co-p~ A    A10    0.975    0.952     11      6 3.40e12 1.06e6  
## 10 10     Conces~ Co-p~ A    A11    0.963    0.881     11      7 1.35e 8 -5.66e4 12  
## # ... with 323 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)
```

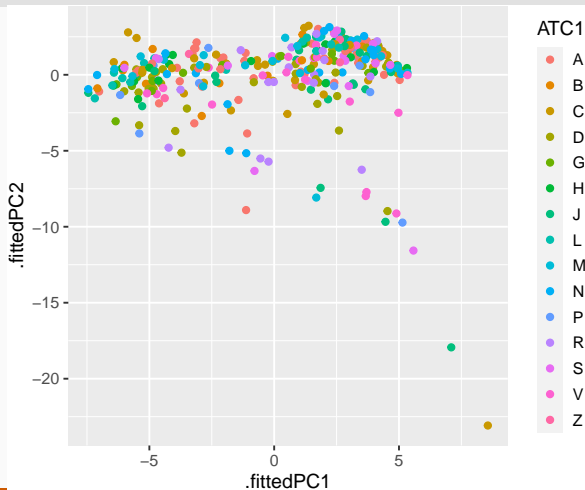
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

```
pcs > ggplot(aes(x=.fittedPC1, y=.fittedPC2, col = ATC1)) +  
  geom_point() + theme(aspect.ratio=1)
```

Principal components  
based on all features  
from the feasts  
package

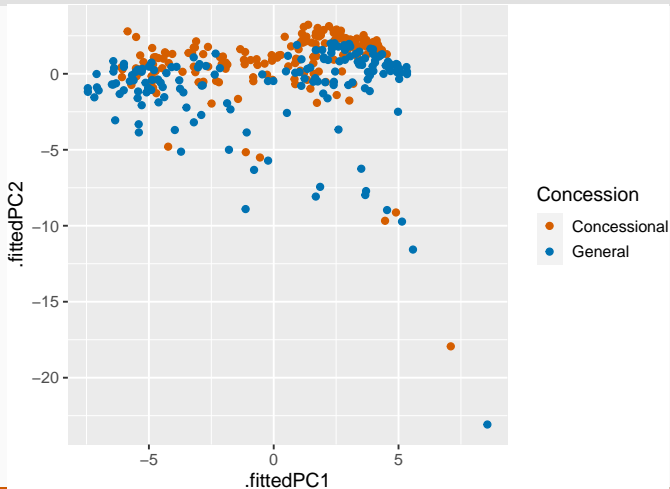




# Feature extraction and statistics

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

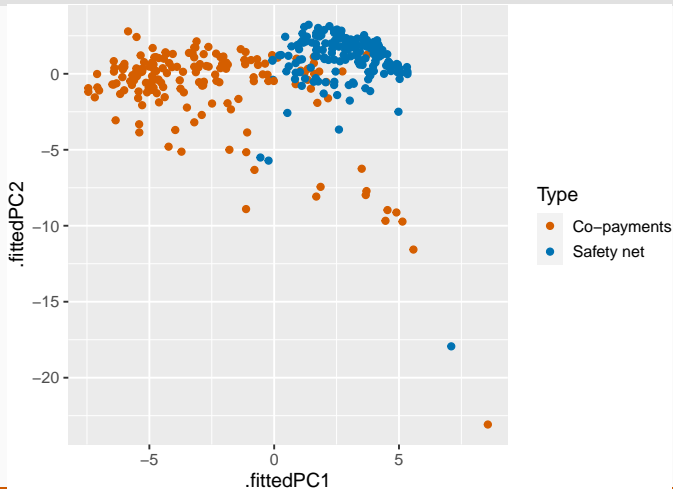
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

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

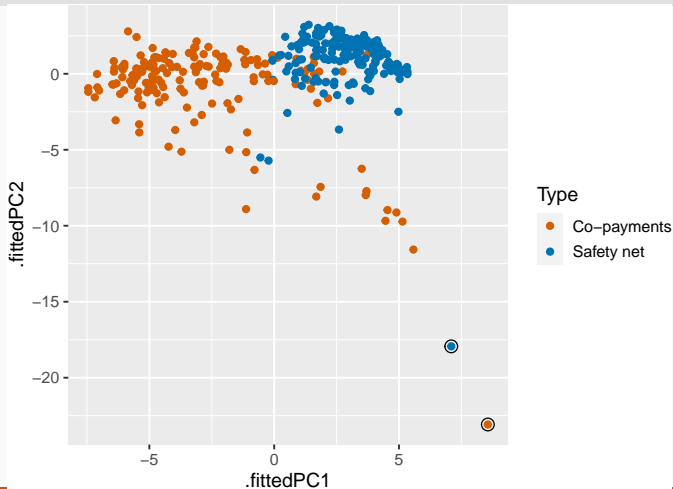
Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

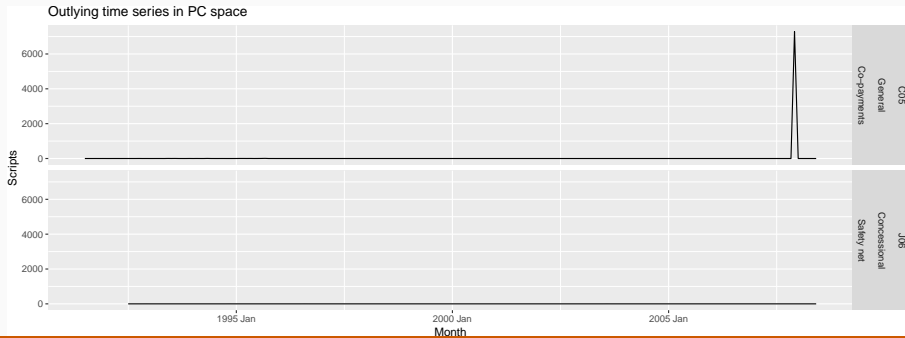
```
pcs > ggplot(aes(x=.fittedPC1, y=.fittedPC2, col=Type)) +  
  geom_point() + theme(aspect.ratio=1)
```

Principal components  
based on all features  
from the feasts  
package



# Feature extraction and statistics

```
outliers <- filter(pcs, .fittedPC2 < -15)
semi_join(PBS, outliers, by = c("ATC1", "ATC2", "Concession", "Type")) ▷
  mutate(Series = glue("{ATC2}", "{Concession}", "{Type}", .sep = "\n\n")) ▷
  ggplot(aes(x = Month, y = Scripts)) +
  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

- Find the most seasonal time series in the tourism data.
- Which state has the strongest trends?
- Use a feature-based approach to look for outlying series in tourism.
- What is unusual about the series you identify as outliers?