Chapter 4 Time Series Features

JUDISMA A. SALI

2023-08-08

4.1 Some simple statistics

```
library(fpp3)
```

```
## -- Attaching packages --------------------------------------------- fpp3 0.5 --
## v tibble      3.2.1     v tsibble     1.1.3
## v dplyr       1.1.2     v tsibbledata 0.4.1
## v tidyr       1.3.0     v feasts      0.3.1
## v lubridate   1.9.2     v fable       0.3.3
## v ggplot2     3.4.2     v fabletools  0.3.3
## -- Conflicts ------------------------------------------------ fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

```
tourism |>
  features(Trips, list(mean = mean)) |>
  arrange(mean)
```

```
## # A tibble: 304 x 4
##    Region          State              Purpose   mean
##    <chr>           <chr>              <chr>    <dbl>
##  1 Kangaroo Island South Australia    Other    0.340
##  2 MacDonnell      Northern Territory Other    0.449
##  3 Wilderness West Tasmania           Other    0.478
##  4 Barkly          Northern Territory Other    0.632
##  5 Clare Valley    South Australia    Other    0.898
##  6 Barossa         South Australia    Other    1.02
##  7 Kakadu Arnhem   Northern Territory Other    1.04
##  8 Lasseter        Northern Territory Other    1.14
##  9 Wimmera         Victoria           Other    1.15
## 10 MacDonnell      Northern Territory Visiting 1.18
## # i 294 more rows
```

```
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 South Australia Business  135.
##  2 1998 Q2 Adelaide South Australia Business  110.
##  3 1998 Q3 Adelaide South Australia Business  166.
##  4 1998 Q4 Adelaide South Australia Business  127.
##  5 1999 Q1 Adelaide South Australia Business  137.
##  6 1999 Q2 Adelaide South Australia Business  200.
##  7 1999 Q3 Adelaide South Australia Business  169.
##  8 1999 Q4 Adelaide South Australia Business  134.
##  9 2000 Q1 Adelaide South Australia Business  154.
## 10 2000 Q2 Adelaide South Australia Business  169.
## # i 24,310 more rows
```

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

```
## # A tibble: 304 x 8
##    Region         State             Purpose    `0%`  `25%`   `50%`  `75%` `100%`
##    <chr>          <chr>             <chr>     <dbl>  <dbl>   <dbl>  <dbl>  <dbl>
##  1 Adelaide       South Australia   Busine~  68.7   134.    153.    177.   242.
##  2 Adelaide       South Australia   Holiday 108.    135.    154.    172.   224.
##  3 Adelaide       South Australia   Other    25.9    43.9   53.8    62.5  107.
##  4 Adelaide       South Australia   Visiti~ 137.    179.    206.    229.   270.
##  5 Adelaide Hills South Australia   Busine~   0       0      1.26    3.92  28.6
##  6 Adelaide Hills South Australia   Holiday   0       5.77   8.52   14.1   35.8
##  7 Adelaide Hills South Australia   Other     0       0      0.908   2.09   8.95
##  8 Adelaide Hills South Australia   Visiti~   0.778   8.91  12.2    16.8   81.1
##  9 Alice Springs  Northern Territo~ Busine~   1.01    9.13  13.3    18.5   34.1
## 10 Alice Springs  Northern Territo~ Holiday   2.81   16.9   31.5    44.8   76.5
## # i 294 more rows
```

4.2 ACF features

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

```
## # A tibble: 304 x 10
##    Region         State Purpose     acf1 acf10 diff1_acf1 diff1_acf10 diff2_acf1
##    <chr>          <chr> <chr>      <dbl> <dbl>      <dbl>       <dbl>      <dbl>
##  1 Adelaide       Sout~ Busine~  0.0333  0.131     -0.520       0.463     -0.676
##  2 Adelaide       Sout~ Holiday  0.0456  0.372     -0.343       0.614     -0.487
##  3 Adelaide       Sout~ Other    0.517   1.15      -0.409       0.383     -0.675
##  4 Adelaide       Sout~ Visiti~  0.0684  0.294     -0.394       0.452     -0.518
##  5 Adelaide Hills Sout~ Busine~  0.0709  0.134     -0.580       0.415     -0.750
##  6 Adelaide Hills Sout~ Holiday  0.131   0.313     -0.536       0.500     -0.716
##  7 Adelaide Hills Sout~ Other    0.261   0.330     -0.253       0.317     -0.457
##  8 Adelaide Hills Sout~ Visiti~  0.139   0.117     -0.472       0.239     -0.626
##  9 Alice Springs  Nort~ Busine~  0.217   0.367     -0.500       0.381     -0.658
## 10 Alice Springs  Nort~ Holiday -0.00660 2.11      -0.153       2.11      -0.274
## # i 294 more rows
## # i 2 more variables: diff2_acf10 <dbl>, season_acf1 <dbl>
```

4.3 STL Features

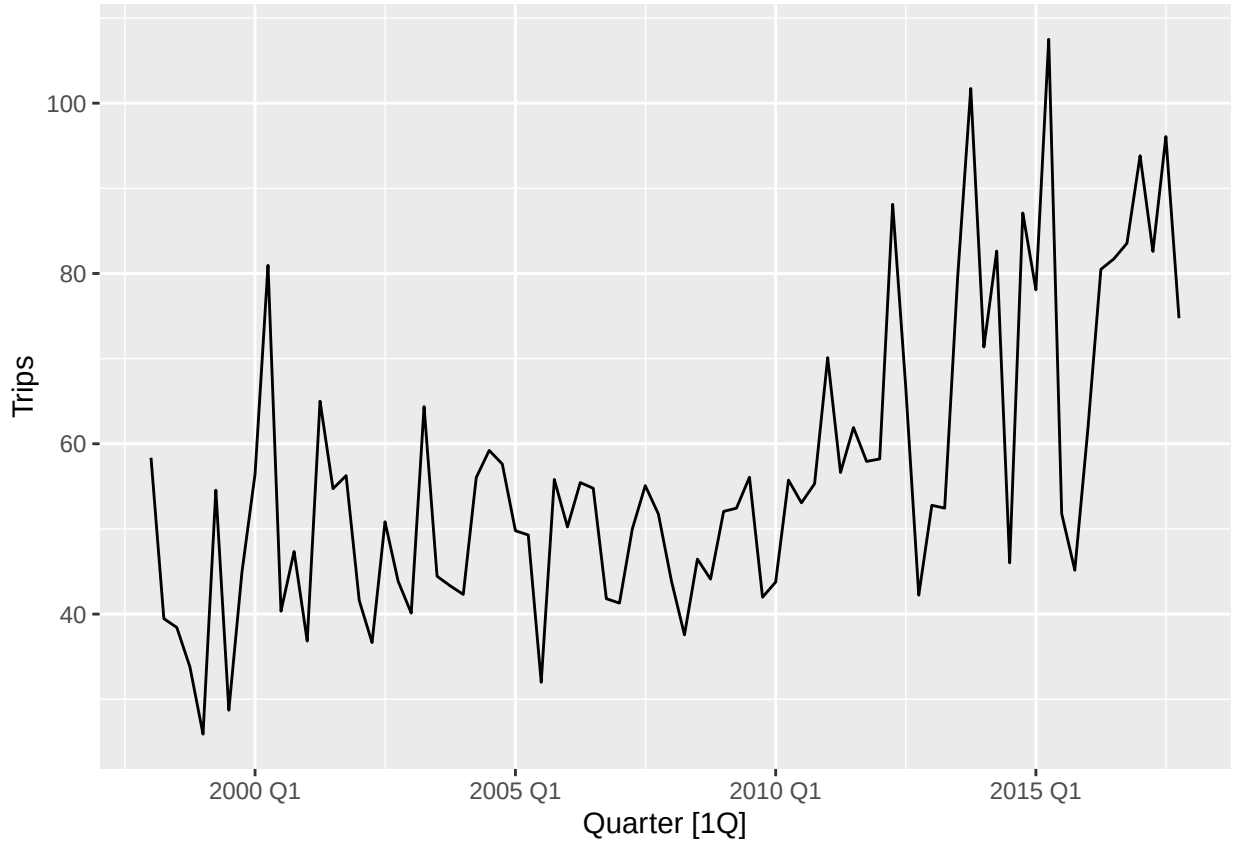
```
tourism |>
  features(Trips, feat_stl)
```

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

```
view(tourism)
```

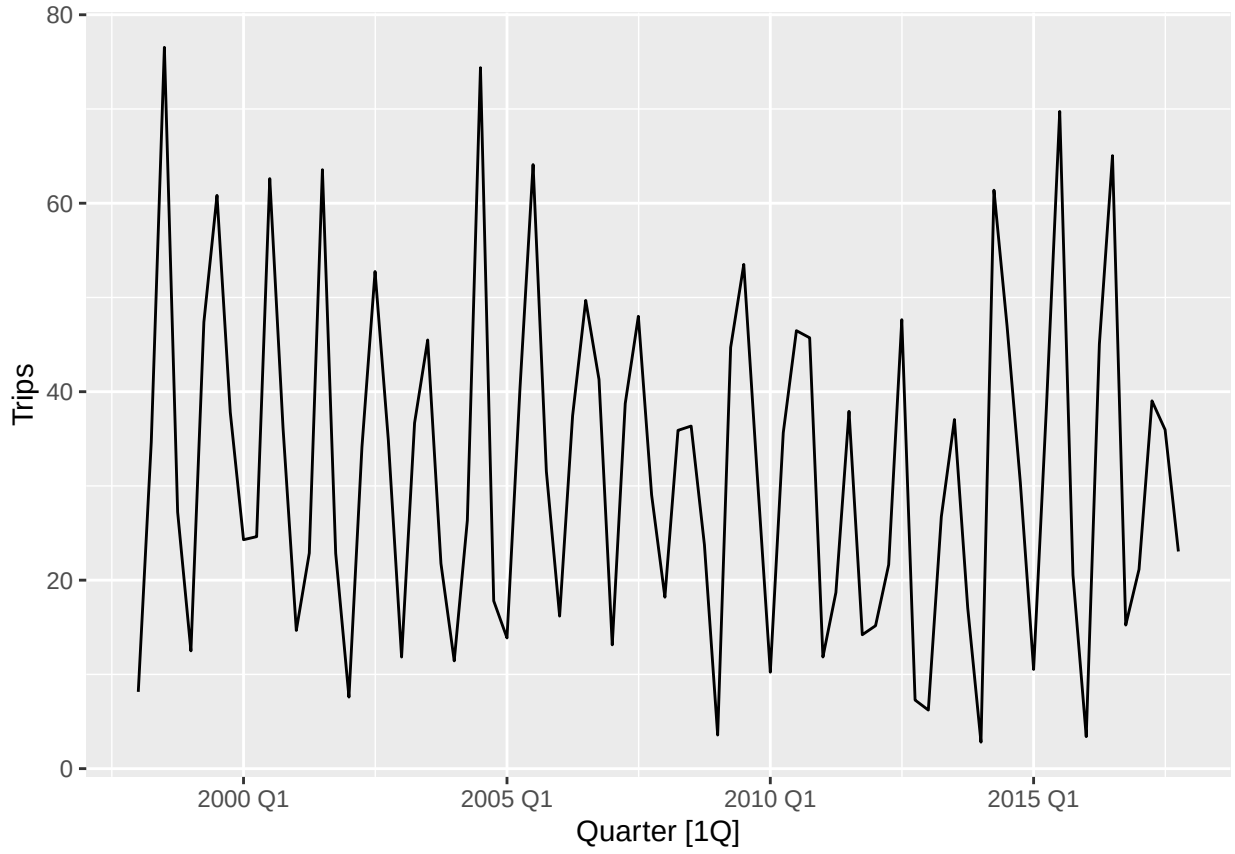
```
tourism |>
  filter(Region=="Adelaide",State=="South Australia",Purpose=="Other")|>

  autoplot(Trips)
```




```
tourism |> filter(Region=="Alice Springs",State=="Northern Territory",Purpose=="Holiday")|>

  autoplot(Trips)
```




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

Most Seasonal

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

```
## # A tibble: 1 x 12
##   Region  State Purpose trend_strength seasonal_strength_year seasonal_peak_year
##   <chr>   <chr> <chr>            <dbl>                  <dbl>              <dbl>
## 1 Snowy ~ New ~ Holiday          0.595                  0.967                  3
## # i 6 more variables: seasonal_trough_year <dbl>, spikiness <dbl>,
## #   linearity <dbl>, curvature <dbl>, stl_e_acf1 <dbl>, stl_e_acf10 <dbl>
```

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

```
## Warning: `n_flat_spots()` was deprecated in feasts 0.1.5.
## i Please use `longest_flat_spot()` instead.
## i The deprecated feature was likely used in the fabletools package.
##   Please report the issue at <https://github.com/tidyverts/fabletools/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: 304 errors (1 unique) encountered for feature 20
```

```
## [304] The `fracdiff` package must be installed to use this functionality. It can be installed with i
```

```
tourism_features
```

```
## # A tibble: 304 x 50
##    Region State Purpose trend_strength seasonal_strength_year seasonal_peak_year
##    <chr>  <chr> <chr>            <dbl>                  <dbl>              <dbl>
##  1 Adela~ Sout~ Busine~          0.464                  0.407                  3
##  2 Adela~ Sout~ Holiday          0.554                  0.619                  1
##  3 Adela~ Sout~ Other            0.746                  0.202                  2
##  4 Adela~ Sout~ Visiti~          0.435                  0.452                  1
##  5 Adela~ Sout~ Busine~          0.464                  0.179                  3
##  6 Adela~ Sout~ Holiday          0.528                  0.296                  2
##  7 Adela~ Sout~ Other            0.593                  0.404                  2
##  8 Adela~ Sout~ Visiti~          0.488                  0.254                  0
##  9 Alice~ Nort~ Busine~          0.534                  0.251                  0
## 10 Alice~ Nort~ Holiday          0.381                  0.832                  3
## # i 294 more rows
## # i 44 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>, diff1_acf10 <dbl>,
## #   diff2_acf1 <dbl>, diff2_acf10 <dbl>, season_acf1 <dbl>, pacf5 <dbl>,
## #   diff1_pacf5 <dbl>, diff2_pacf5 <dbl>, season_pacf <dbl>,
## #   zero_run_mean <dbl>, nonzero_squared_cv <dbl>, zero_start_prop <dbl>, ...
```

4.5 Exploring Australian tourism data

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

```
## Warning: 304 errors (1 unique) encountered for feature 20
## [304] The `fracdiff` package must be installed to use this functionality. It can be installed with i
```

```
tourism_features
```

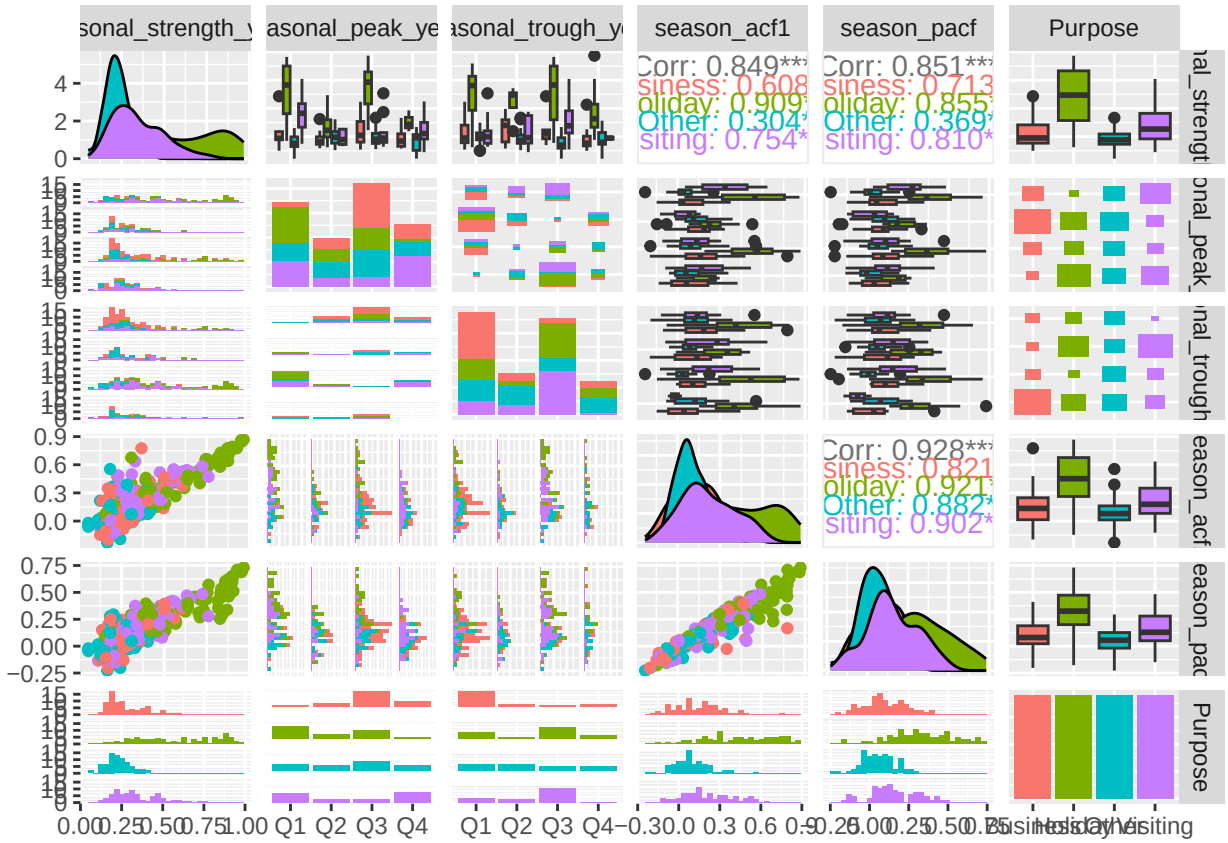
```
## # A tibble: 304 x 50
##    Region State Purpose trend_strength seasonal_strength_year seasonal_peak_year
##    <chr>  <chr> <chr>            <dbl>                  <dbl>              <dbl>
##  1 Adela~ Sout~ Busine~          0.464                  0.407                  3
##  2 Adela~ Sout~ Holiday          0.554                  0.619                  1
##  3 Adela~ Sout~ Other            0.746                  0.202                  2
##  4 Adela~ Sout~ Visiti~          0.435                  0.452                  1
##  5 Adela~ Sout~ Busine~          0.464                  0.179                  3
##  6 Adela~ Sout~ Holiday          0.528                  0.296                  2
##  7 Adela~ Sout~ Other            0.593                  0.404                  2
##  8 Adela~ Sout~ Visiti~          0.488                  0.254                  0
##  9 Alice~ Nort~ Busine~          0.534                  0.251                  0
## 10 Alice~ Nort~ Holiday          0.381                  0.832                  3
## # i 294 more rows
## # i 44 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>, diff1_acf10 <dbl>,
## #   diff2_acf1 <dbl>, diff2_acf10 <dbl>, season_acf1 <dbl>, pacf5 <dbl>,
## #   diff1_pacf5 <dbl>, diff2_pacf5 <dbl>, season_pacf <dbl>,
## #   zero_run_mean <dbl>, nonzero_squared_cv <dbl>, zero_start_prop <dbl>, ...
```

```
library(glue)
```

```
tourism_features |>
  select_at(vars(contains("season"), Purpose)) |>
  mutate(
    seasonal_peak_year = seasonal_peak_year +
      4*(seasonal_peak_year==0),
    seasonal_trough_year = seasonal_trough_year +
      4*(seasonal_trough_year==0),
    seasonal_peak_year = glue("Q{seasonal_peak_year}"),
    seasonal_trough_year = glue("Q{seasonal_trough_year}"),
  ) |>
  GGally::ggpairs(mapping = aes(colour = Purpose))
```

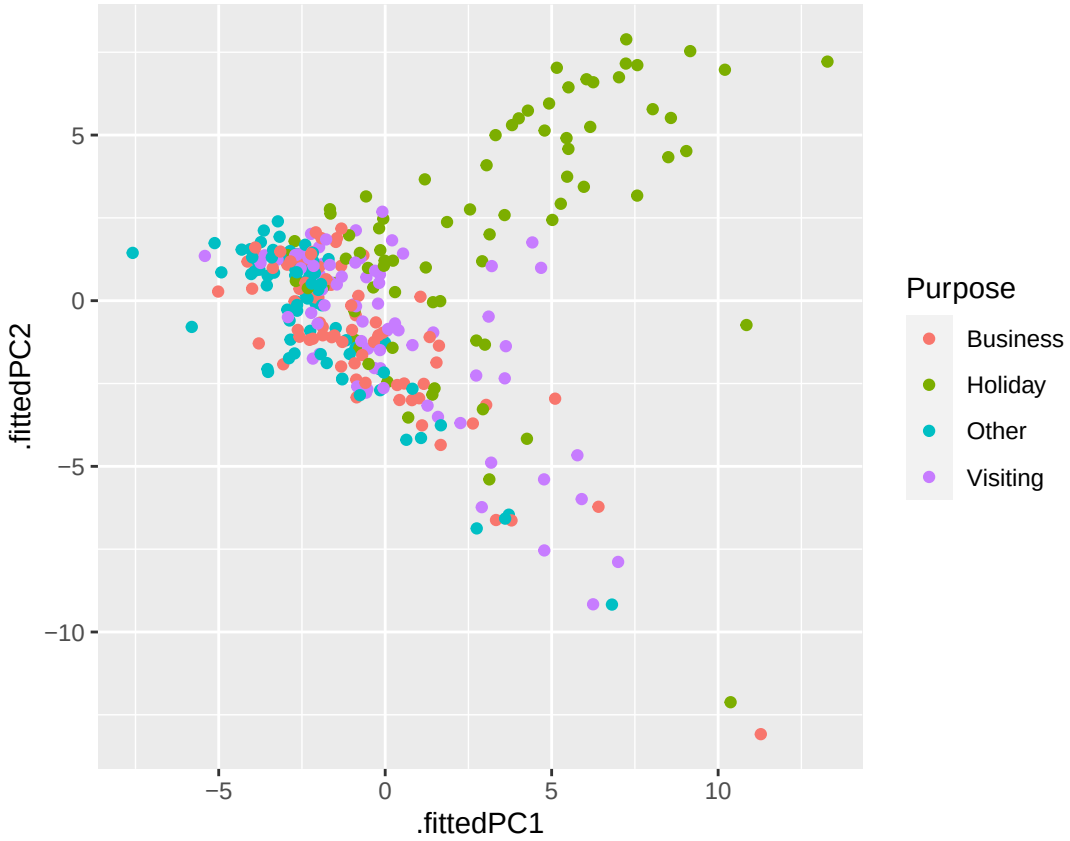
```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
library(broom)
```

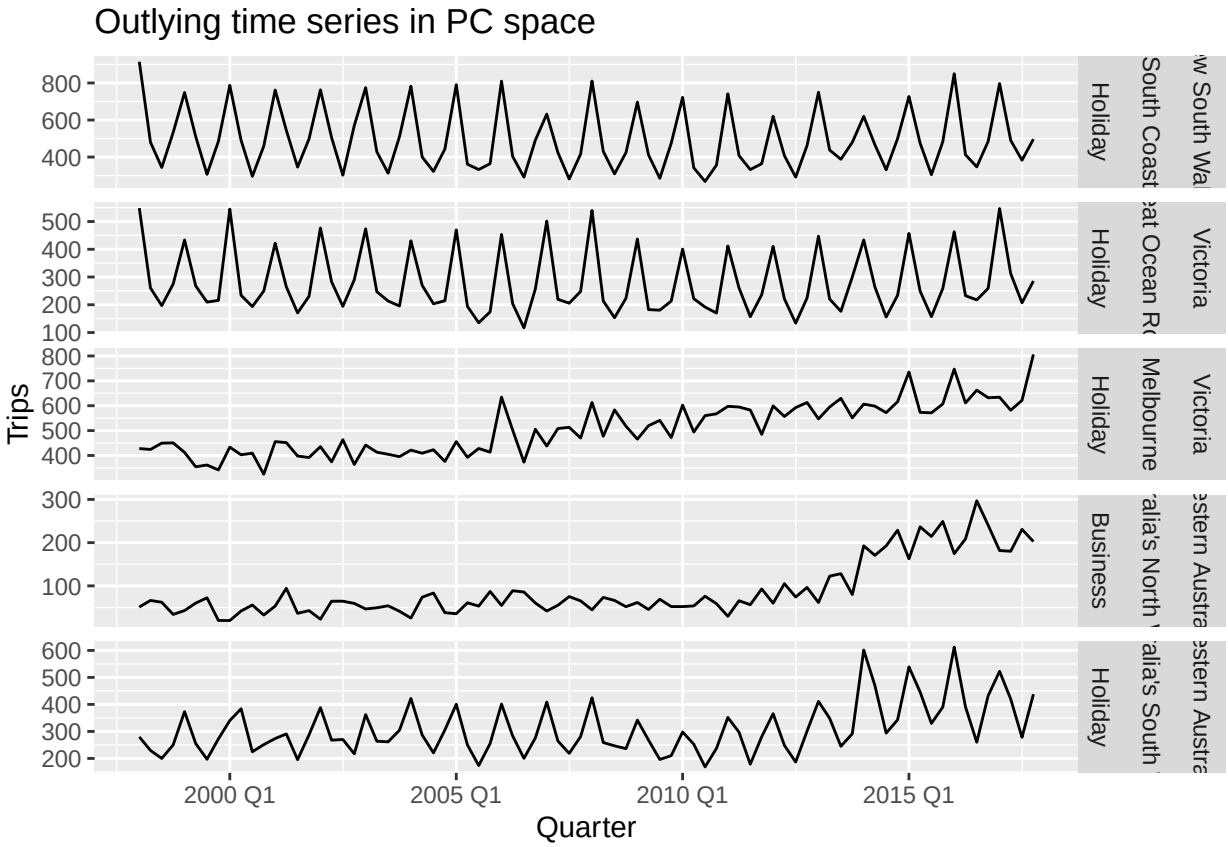
```
pcs <- tourism_features |>
  select(-State, -Region, -Purpose) |>
  prcomp(scale = TRUE) |>
  augment(tourism_features)
pcs |>
  ggplot(aes(x = .fittedPC1, y = .fittedPC2, col = Purpose)) +
  geom_point() +
  theme(aspect.ratio = 1)
```




```
outliers <- pcs |>
  filter(.fittedPC1 > 10) |>
  select(Region, State, Purpose, .fittedPC1, .fittedPC2)
outliers
```

```
## # A tibble: 5 x 5
##   Region                 State             Purpose  .fittedPC1 .fittedPC2
##   <chr>                  <chr>             <chr>         <dbl>      <dbl>
## 1 Australia's North West Western Australia Business       11.3    -13.1
## 2 Australia's South West Western Australia Holiday        10.9     -0.734
## 3 Great Ocean Road       Victoria          Holiday        10.2      6.97
## 4 Melbourne              Victoria          Holiday        10.4    -12.1
## 5 South Coast            New South Wales   Holiday        13.3      7.22
```

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


Outlying time series in PC space
Trips
Quarter
2000 Q1
2005 Q1
2010 Q1
2015 Q1
Holiday
Holiday
Holiday
Business
Holiday