Feasts & fables

Modern tools for time series analysis

Rob J Hyndman

robjhyndman.com/whyR2021

MONASH University

Outline

Tidyverts packages

tidyverts.org

tsibble

tsibbledata

feasts

fable

Outline

Australian electricity demand

```
## # A tsibble: 420,864 x 6 [30m] <Australia/Melbourne>
## # Key:       State [8]
##    Time                State Date       Holiday Temperature Demand
##    <dttm>              <fct> <date>     <lgl>         <dbl>  <dbl>
##  1 2012-01-01 00:00:00 VIC   2012-01-01 TRUE           21.4  4383.
##  2 2012-01-01 00:30:00 VIC   2012-01-01 TRUE           21.0  4263.
##  3 2012-01-01 01:00:00 VIC   2012-01-01 TRUE           20.7  4049.
##  4 2012-01-01 01:30:00 VIC   2012-01-01 TRUE           20.6  3878.
##  5 2012-01-01 02:00:00 VIC   2012-01-01 TRUE           20.4  4036.
##  6 2012-01-01 02:30:00 VIC   2012-01-01 TRUE           20.2  3866.
##  7 2012-01-01 03:00:00 VIC   2012-01-01 TRUE           20.1  3694.
##  8 2012-01-01 03:30:00 VIC   2012-01-01 TRUE           19.6  3562.
##  9 2012-01-01 04:00:00 VIC   2012-01-01 TRUE           19.1  3433.
## 10 2012-01-01 04:30:00 VIC   2012-01-01 TRUE           19.0  3359.
## # ... with 420,854 more rows
```

Australian electricity demand

```
## # A tsibble: 420,864 x 6 [30m] <Australia/Melbourne>
## # Key:       State [8]
##    Time                State Date       Holiday Temperature Demand
##                        <fct> <date>     <lgl>         <dbl>  <dbl>
##  1 2012-01-01 00:00:00 VIC   2012-01-01 TRUE           21.4  4383.
##  2 2012-01-01 00:30:00 VIC   2012-01-01 TRUE           21.0  4263.
##  3 2012-01-01 01:00:00 VIC   2012-01-01 TRUE           20.7  4049.
##  4 2012-01-01 01:30:00 VIC   2012-01-01 TRUE           20.6  3878.
##  5 2012-01-01 02:00:00 VIC   2012-01-01 TRUE           20.4  4036.
##  6 2012-01-01 02:30:00 VIC   2012-01-01 TRUE           20.2  3866.
##  7 2012-01-01 03:00:00 VIC   2012-01-01 TRUE           20.1  3694.
##  8 2012-01-01 03:30:00 VIC   2012-01-01 TRUE           19.6  3562.
##  9 2012-01-01 04:00:00 VIC   2012-01-01 TRUE           19.1  3433.
## 10 2012-01-01 04:30:00 VIC   2012-01-01 TRUE           19.0  3359.
## # ... with 420,854 more rows
```

Index

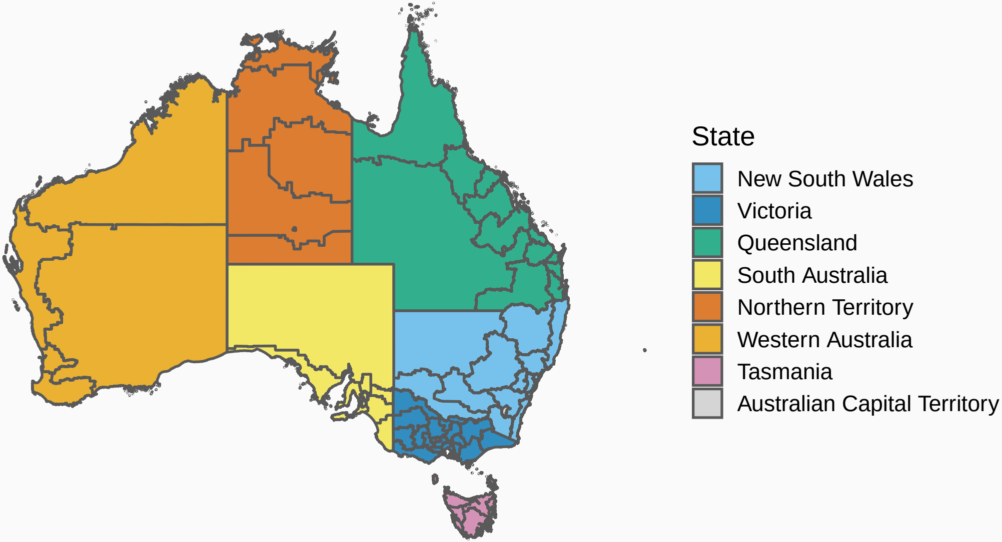
Australian electricity demand

```
## # A tsibble: 420,864 x 6 [30m] <Australia/Melbourne>
## # Key:       State [8]
##    Time                State Date       Holiday Temperature Demand
##    Index               Key   <date>     <lgl>         <dbl>  <dbl>
##  1 2012-01-01 00:00:00 VIC   2012-01-01 TRUE           21.4  4383.
##  2 2012-01-01 00:30:00 VIC   2012-01-01 TRUE           21.0  4263.
##  3 2012-01-01 01:00:00 VIC   2012-01-01 TRUE           20.7  4049.
##  4 2012-01-01 01:30:00 VIC   2012-01-01 TRUE           20.6  3878.
##  5 2012-01-01 02:00:00 VIC   2012-01-01 TRUE           20.4  4036.
##  6 2012-01-01 02:30:00 VIC   2012-01-01 TRUE           20.2  3866.
##  7 2012-01-01 03:00:00 VIC   2012-01-01 TRUE           20.1  3694.
##  8 2012-01-01 03:30:00 VIC   2012-01-01 TRUE           19.6  3562.
##  9 2012-01-01 04:00:00 VIC   2012-01-01 TRUE           19.1  3433.
## 10 2012-01-01 04:30:00 VIC   2012-01-01 TRUE           19.0  3359.
## # ... with 420,854 more rows
```

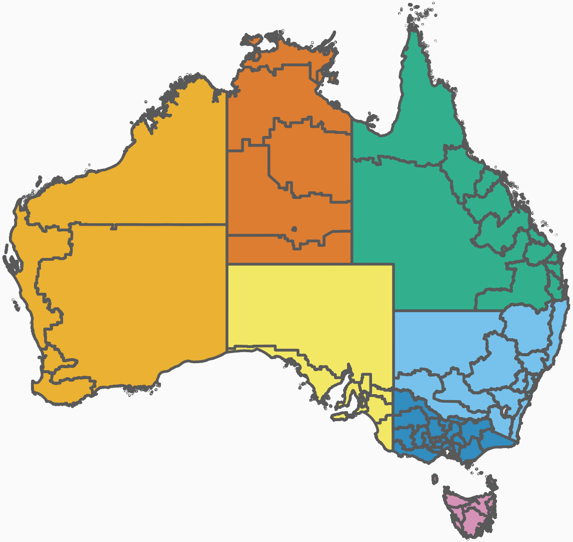
Australian electricity demand

```
## # A tsibble: 420,864 x 6 [30m] <Australia/Melbourne>
## # Key:       State [8]
##    Time                State Date       Holiday Temperature Demand
##     Index              Key   Measures
##  1 2012-01-01 00:00:00 VIC   2012-01-01 TRUE           21.4  4383.
##  2 2012-01-01 00:30:00 VIC   2012-01-01 TRUE           21.0  4263.
##  3 2012-01-01 01:00:00 VIC   2012-01-01 TRUE           20.7  4049.
##  4 2012-01-01 01:30:00 VIC   2012-01-01 TRUE           20.6  3878.
##  5 2012-01-01 02:00:00 VIC   2012-01-01 TRUE           20.4  4036.
##  6 2012-01-01 02:30:00 VIC   2012-01-01 TRUE           20.2  3866.
##  7 2012-01-01 03:00:00 VIC   2012-01-01 TRUE           20.1  3694.
##  8 2012-01-01 03:30:00 VIC   2012-01-01 TRUE           19.6  3562.
##  9 2012-01-01 04:00:00 VIC   2012-01-01 TRUE           19.1  3433.
## 10 2012-01-01 04:30:00 VIC   2012-01-01 TRUE           19.0  3359.
## # ... with 420,854 more rows
```

Australian tourism regions



Australian tourism regions



- Quarterly data on visitor nights: 1998 – 2017
- From *National Visitor Survey*, interviews of 120,000 Australians aged 15+.
- Geographical hierarchy split by
  - 8 states and territories
  - 76 regions
- Purpose:
  - Holidays
  - Business
  - Visiting friends & relatives
  - Other

Quarterly tourism data

```
## # 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
```

Domestic visitor nights in thousands by state/region and purpose.

Quarterly tourism data

```
## # 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
```

Domestic visitor nights in thousands by state/region and purpose.

Quarterly tourism data

```
## # 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
```

Domestic visitor nights in thousands by state/region and purpose.

Quarterly tourism data

```
## # 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
```

Domestic visitor nights in thousands by state/region and purpose.

Characteristics of modern time series

- Often observed at sub-daily frequency over a long time.
- Multiple keys which may be nested.
- Multiple seasonal patterns.
- Multiple measures for each combination of index and keys.

Characteristics of modern time series

- Often observed at sub-daily frequency over a long time.
- Multiple keys which may be nested.
- Multiple seasonal patterns.
- Multiple measures for each combination of index and keys.

`tsibble` objects

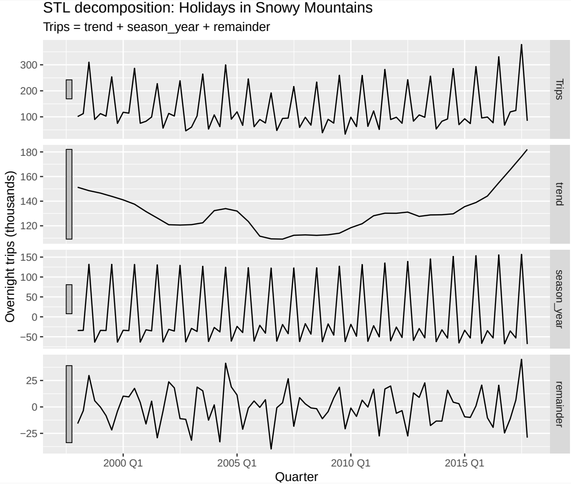
- A `tsibble` allows storage and manipulation of multiple time series in R.
- It contains:
  - An index: time information about the observation
  - Key variable(s): optional unique identifiers for each series
  - Measured variable(s): numbers of interest and any other variable

Outline

STL decomposition

STL decomposition

$$y_t = T_t + S_t + R_t$$



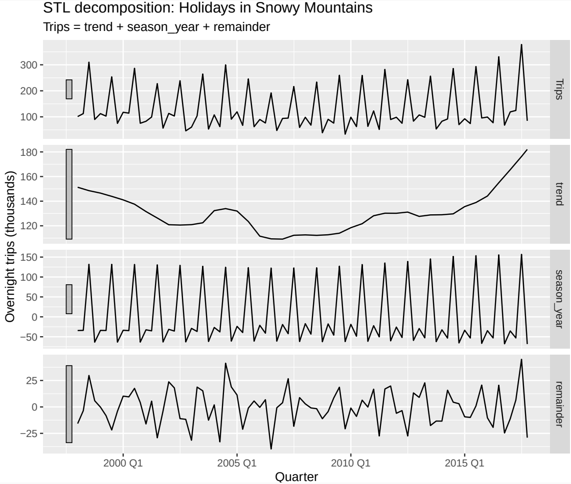
y_t

T_t

S_t

R_t

STL decomposition



y_t

T_t

S_t

R_t

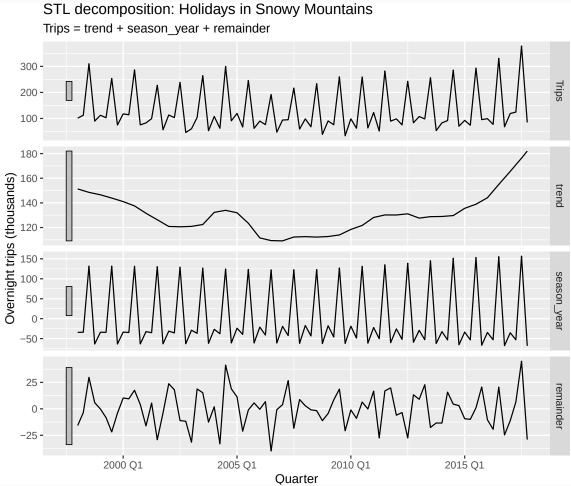
STL decomposition

$$y_t = T_t + S_t + R_t$$

Trend strength

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

STL decomposition



y_t

T_t

S_t

R_t

STL decomposition

$$y_t = T_t + S_t + R_t$$

Trend strength

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

Seasonal strength

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

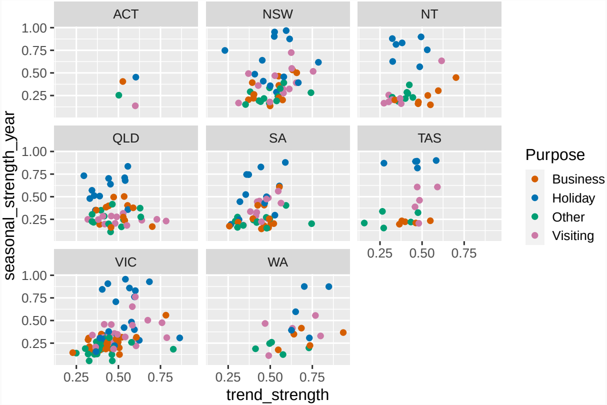
STL-based features

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

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

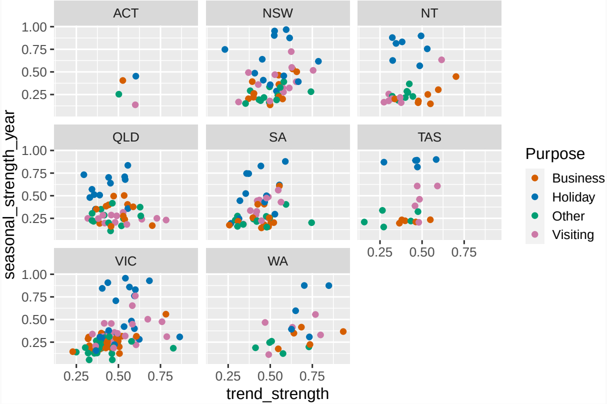
STL-based features

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



STL-based features

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

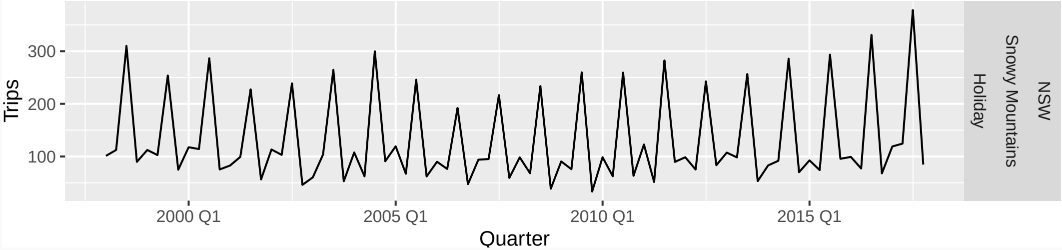


- Holidays more seasonal than other travel.
- WA has strongest trends.

STL-based features

Find the most seasonal time series:

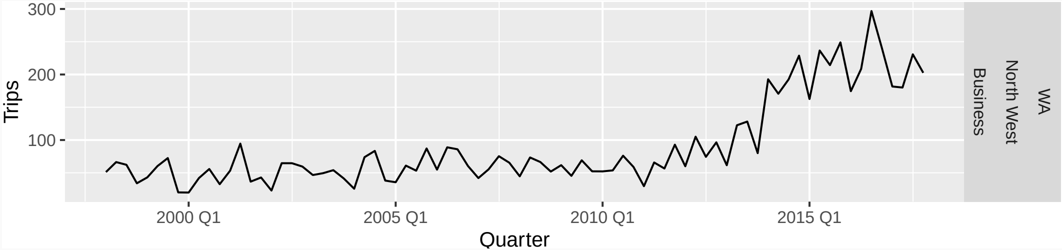
```
tourism %>%
  features(Trips, feat_stl) %>%
  filter(seasonal_strength_year == max(seasonal_strength_year)) %>%
  left_join(tourism, by = c("State", "Region", "Purpose")) %>%
  ggplot(aes(x = Quarter, y = Trips)) +
  geom_line() +
  facet_grid(vars(State, Region, Purpose))
```



STL-based features

Find the most trended time series:

```
tourism %>%
  features(Trips, feat_stl) %>%
  filter(trend_strength == max(trend_strength)) %>%
  left_join(tourism, by = c("State", "Region", "Purpose")) %>%
  ggplot(aes(x = Quarter, y = Trips)) +
  geom_line() +
  facet_grid(vars(State, Region, Purpose))
```



Time series features

```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_strength seasonal_streng~ seasonal_peak_y~ seasonal_trough~
##    <chr>   <chr> <chr>            <dbl>            <dbl>            <dbl>            <dbl>
##  1 Adelai~ SA    Busine~          0.464            0.407                3                1
##  2 Adelai~ SA    Holiday          0.554            0.619                1                2
##  3 Adelai~ SA    Other            0.746            0.202                2                1
##  4 Adelai~ SA    Visiti~          0.435            0.452                1                3
##  5 Adelai~ SA    Busine~          0.464            0.179                3                0
##  6 Adelai~ SA    Holiday          0.528            0.296                2                1
##  7 Adelai~ SA    Other            0.593            0.404                2                2
##  8 Adelai~ SA    Visiti~          0.488            0.254                0                3
##  9 Alice ~ NT    Busine~          0.534            0.251                0                1
## 10 Alice ~ NT    Holiday          0.381            0.832                3                1
## # ... with 294 more rows, and 44 more variables: 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>, zero_end_prop <dbl>, lambda_guerrero <dbl>, kpss_stat <dbl>,
## #   kpss_pvalue <dbl>, pp_stat <dbl>, pp_pvalue <dbl>, ndiffs <int>, nsdiffs <int>, ...
```

Reduced feature space

```r
pcs <- tourism_features %>%
  select(-State, -Region, -Purpose) %>%
  prcomp(scale = TRUE) %>%
  augment(tourism_features)
```

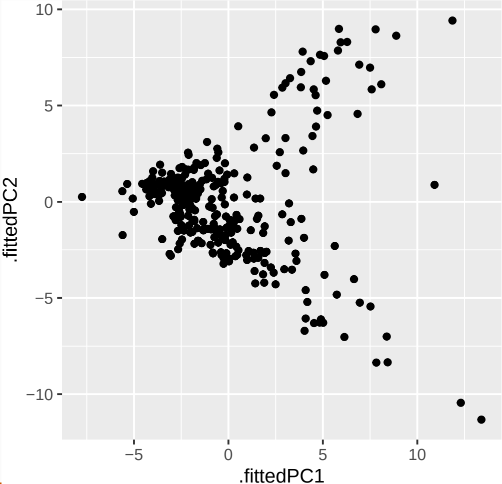
Principal components based on all features from the feasts package

```
## # A tibble: 304 x 100
##    .rownames Region         State Purpose trend_strength seasonal_streng~ seasonal_peak_y~
##    <chr>     <chr>          <chr> <chr>            <dbl>            <dbl>            <dbl>
##  1 1         Adelaide       SA    Busine~          0.464            0.407                3
##  2 2         Adelaide       SA    Holiday          0.554            0.619                1
##  3 3         Adelaide       SA    Other            0.746            0.202                2
##  4 4         Adelaide       SA    Visiti~          0.435            0.452                1
##  5 5         Adelaide Hills SA    Busine~          0.464            0.179                3
##  6 6         Adelaide Hills SA    Holiday          0.528            0.296                2
##  7 7         Adelaide Hills SA    Other            0.593            0.404                2
##  8 8         Adelaide Hills SA    Visiti~          0.488            0.254                0
##  9 9         Alice Springs  NT    Busine~          0.534            0.251                0
## 10 10        Alice Springs  NT    Holiday          0.381            0.832                3
## # ... with 294 more rows, and 93 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>,
```

Reduced feature space

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

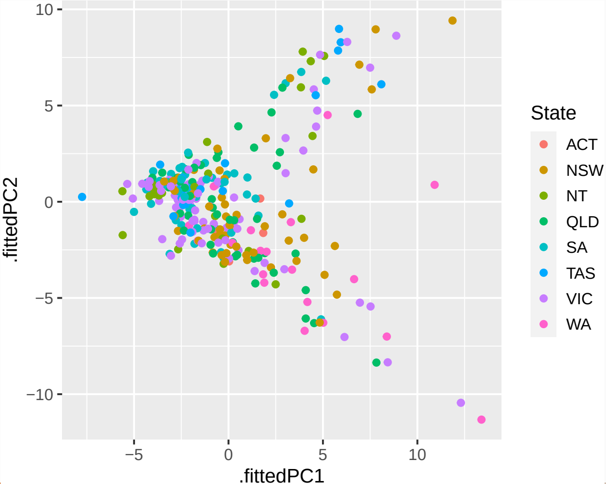
Principal components based on all features from the feasts package



Reduced feature space

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

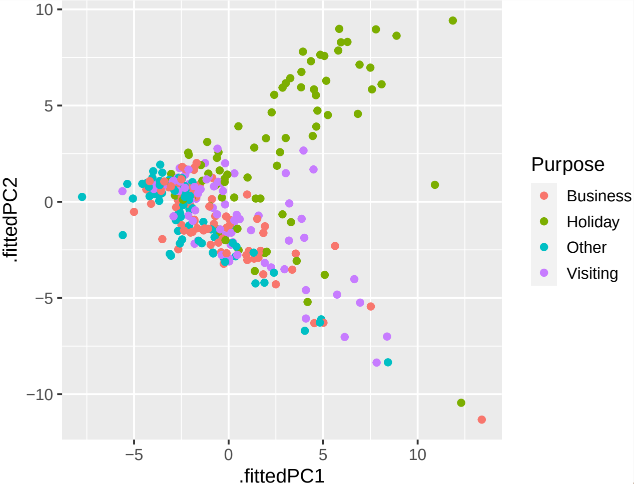
Principal components based on all features from the feasts package



Reduced feature space

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

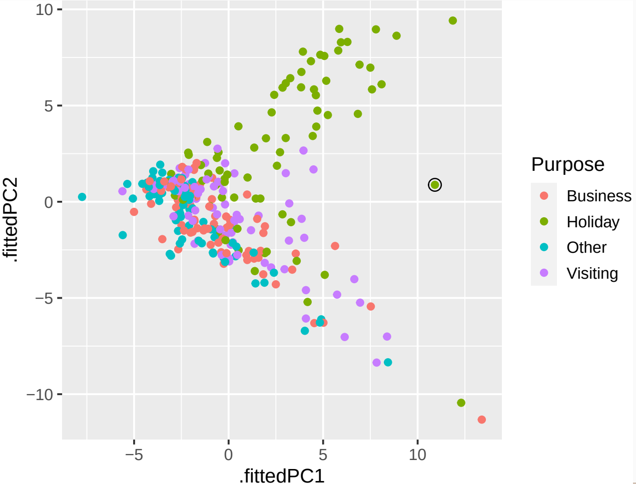
Principal components based on all features from the feasts package



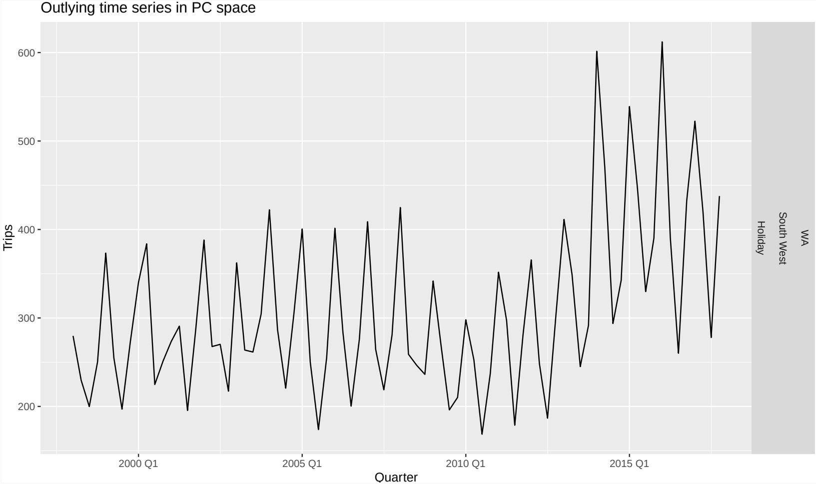
Anomaly detection using time series features

```
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



Anomaly detection using time series features



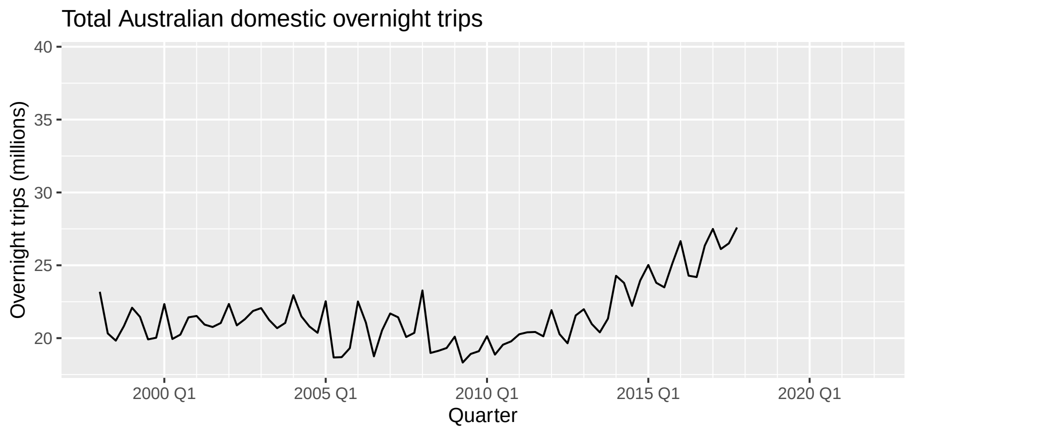
Outline

Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.

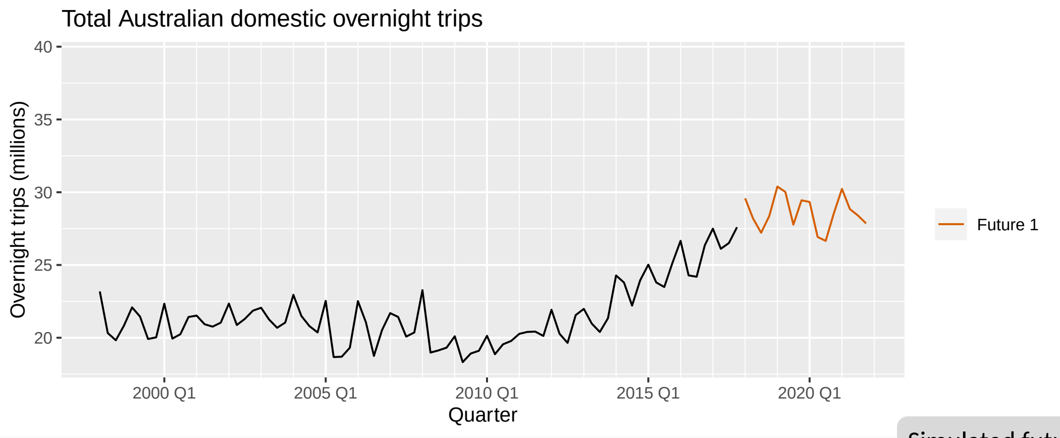
Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Random futures

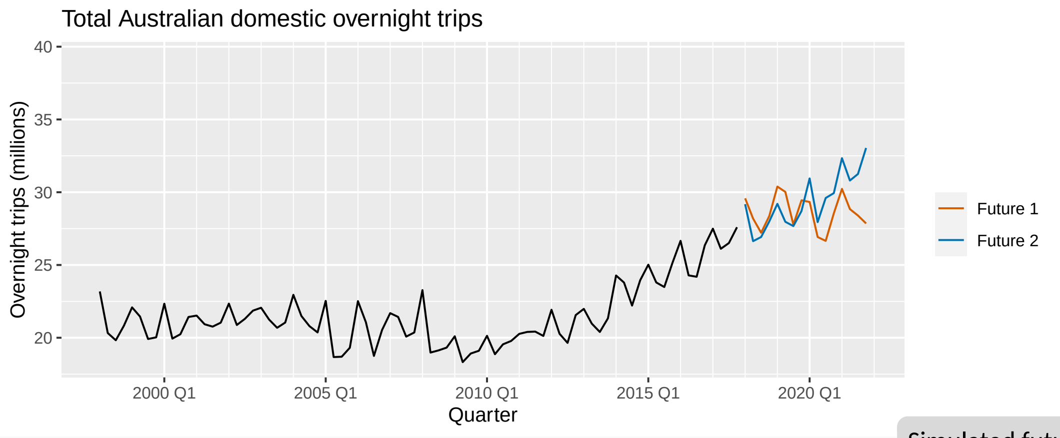
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

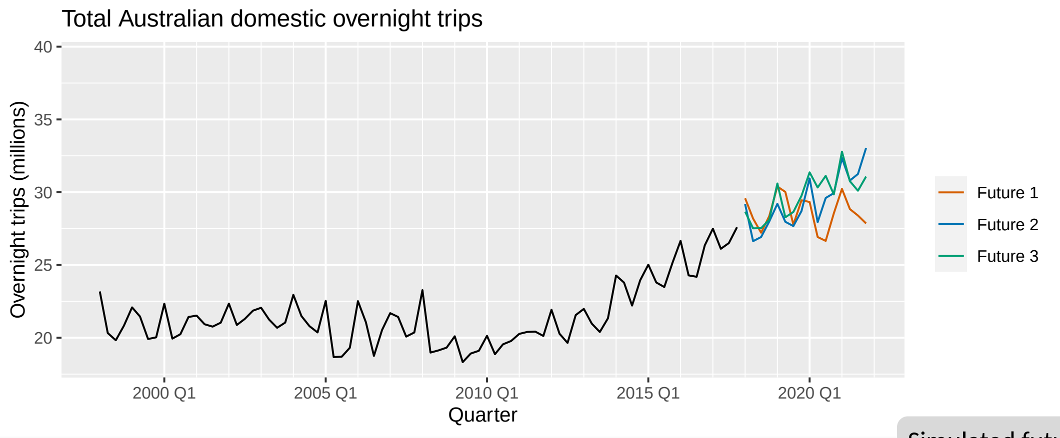
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

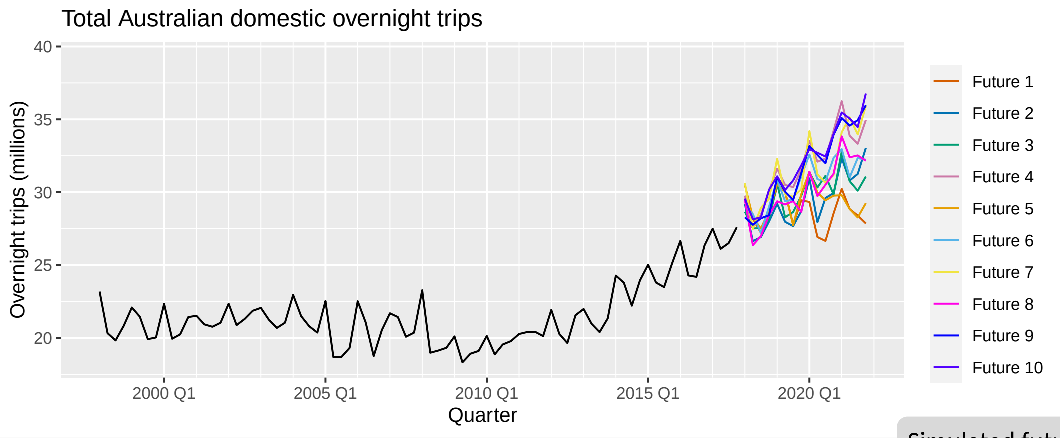
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

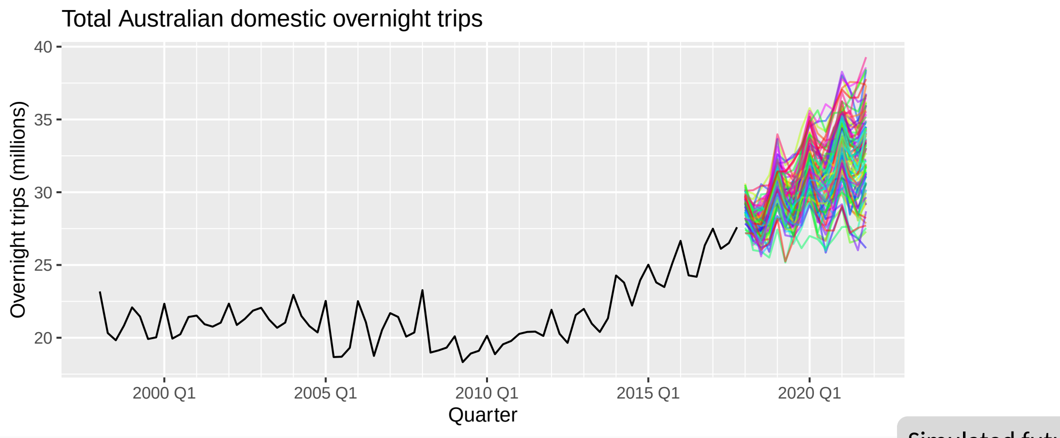
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

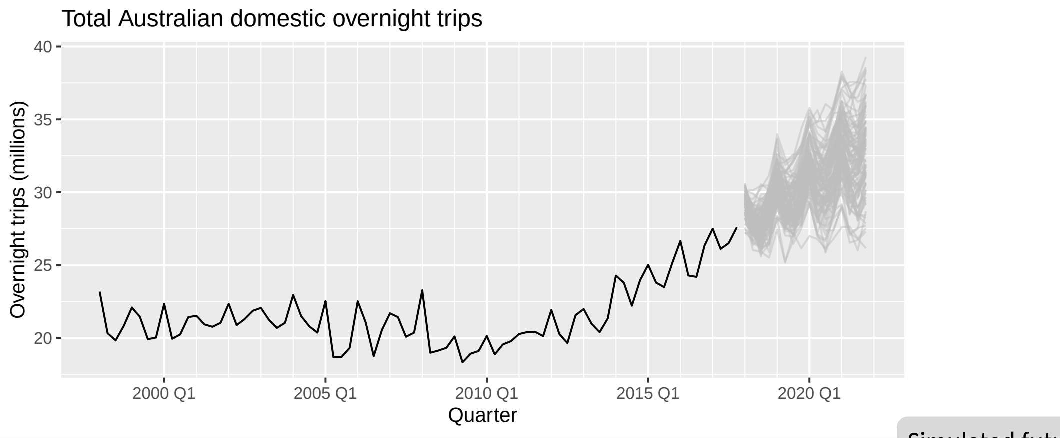
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

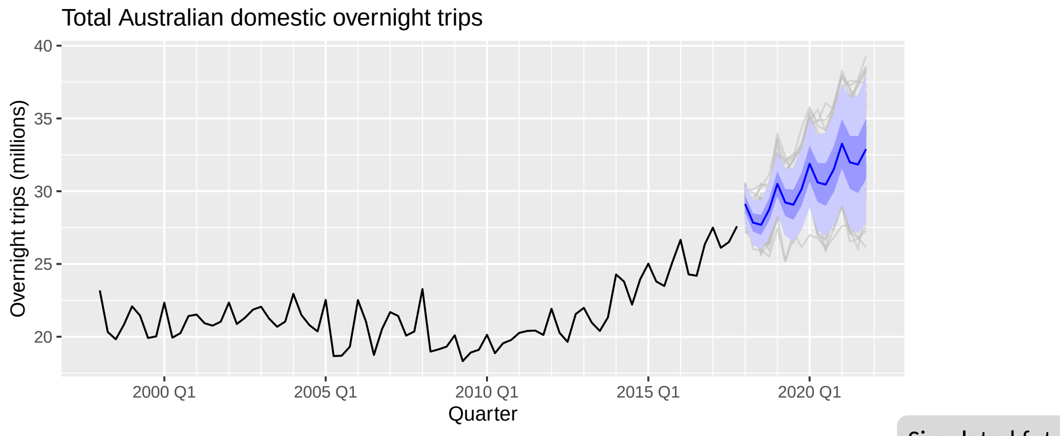
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

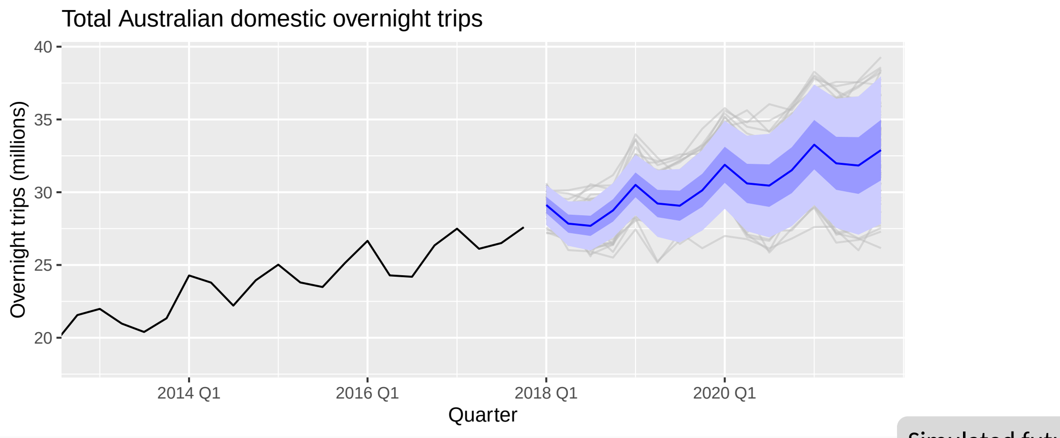
A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



Simulated futures from an ETS model

Model fitting

```r
tourism_fit <- tourism %>%
  filter(year(Quarter) <= 2015) %>%
  model(ets = ETS(Trips), arima = ARIMA(Trips)) %>%
  mutate(ensemble = (ets + arima)/2)
```

```
## # A mable: 304 x 6
## # Key:     Region, State, Purpose [304]
##    Region         State Purpose         ets                              arima      ensemble
##    <chr>          <chr> <chr>       <model>                            <model>       <model>
##  1 Adelaide       SA    Busine~ <ETS(M,N,A)> <ARIMA(0,0,0)(1,0,1)[4] w/ mean> <COMBINATION>
##  2 Adelaide       SA    Holiday <ETS(M,N,A)> <ARIMA(0,0,0)(2,0,0)[4] w/ mean> <COMBINATION>
##  3 Adelaide       SA    Other   <ETS(M,A,N)>          <ARIMA(0,1,1) w/ drift> <COMBINATION>
##  4 Adelaide       SA    Visiti~ <ETS(A,N,A)> <ARIMA(0,0,0)(1,0,1)[4] w/ mean> <COMBINATION>
##  5 Adelaide Hil~ SA    Busine~ <ETS(A,N,N)>           <ARIMA(0,0,0) w/ mean> <COMBINATION>
##  6 Adelaide Hil~ SA    Holiday <ETS(A,A,N)>           <ARIMA(0,0,0) w/ mean> <COMBINATION>
##  7 Adelaide Hil~ SA    Other   <ETS(A,N,N)>         <ARIMA(2,1,1)(2,0,0)[4]> <COMBINATION>
##  8 Adelaide Hil~ SA    Visiti~ <ETS(M,A,A)>                   <ARIMA(0,1,1)> <COMBINATION>
##  9 Alice Springs NT    Busine~ <ETS(A,N,N)> <ARIMA(0,0,0)(0,0,1)[4] w/ mean> <COMBINATION>
## 10 Alice Springs NT    Holiday <ETS(M,N,A)>         <ARIMA(0,0,0)(0,1,2)[4]> <COMBINATION>
## # ... with 294 more rows
```

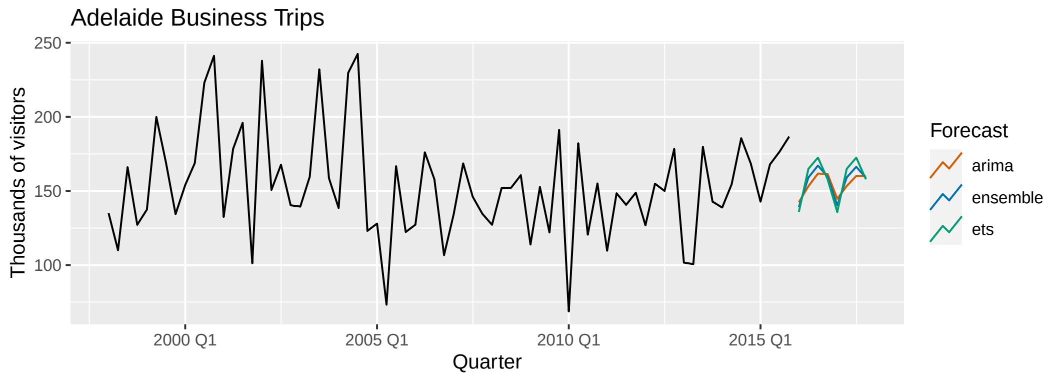
Producing forecasts

```r
tourism_fc <- tourism_fit %>%
  forecast(h = "2 years")
```

```
## # A fable: 7,296 x 7 [1Q]
## # Key:     Region, State, Purpose, .model [912]
##    Region   State Purpose  .model Quarter        Trips .mean
##    <chr>    <chr> <chr>    <chr>    <qtr>       <dist> <dbl>
##  1 Adelaide SA    Business ets    2016 Q1  N(136, 902)  136.
##  2 Adelaide SA    Business ets    2016 Q2 N(165, 1344)  165.
##  3 Adelaide SA    Business ets    2016 Q3 N(173, 1490)  173.
##  4 Adelaide SA    Business ets    2016 Q4 N(158, 1277)  158.
##  5 Adelaide SA    Business ets    2017 Q1  N(136, 979)  136.
##  6 Adelaide SA    Business ets    2017 Q2 N(165, 1422)  165.
##  7 Adelaide SA    Business ets    2017 Q3 N(173, 1569)  173.
##  8 Adelaide SA    Business ets    2017 Q4 N(158, 1356)  158.
##  9 Adelaide SA    Business arima  2016 Q1 N(142, 1232)  142.
## 10 Adelaide SA    Business arima  2016 Q2 N(153, 1232)  153.
## # ... with 7,286 more rows
```

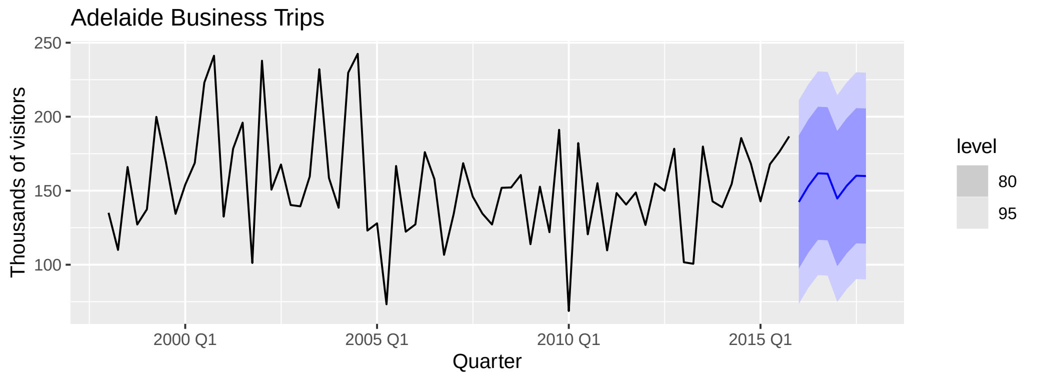
Visualising forecasts

```
tourism_fc %>%
  filter(Region == "Adelaide", Purpose=="Business") %>%
  autoplot(tourism, level = NULL) +
  labs(title = "Adelaide Business Trips", y = "Thousands of visitors") +
  guides(color = guide_legend(title = "Forecast"))
```



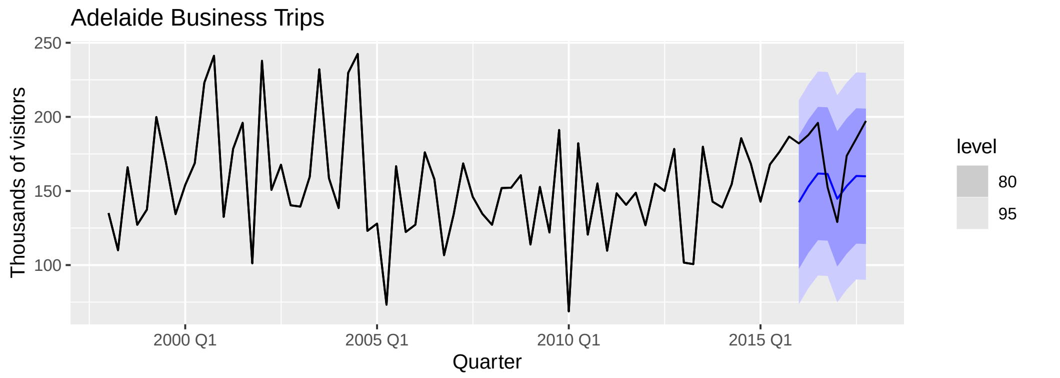
Visualising forecasts

```
tourism_fc %>%
  filter(Region == "Adelaide", Purpose=="Business", .model == "arima") %>%
  autoplot(tourism) +
  labs(title = "Adelaide Business Trips", y = "Thousands of visitors") +
  guides(color = guide_legend(title = "Forecast"))
```



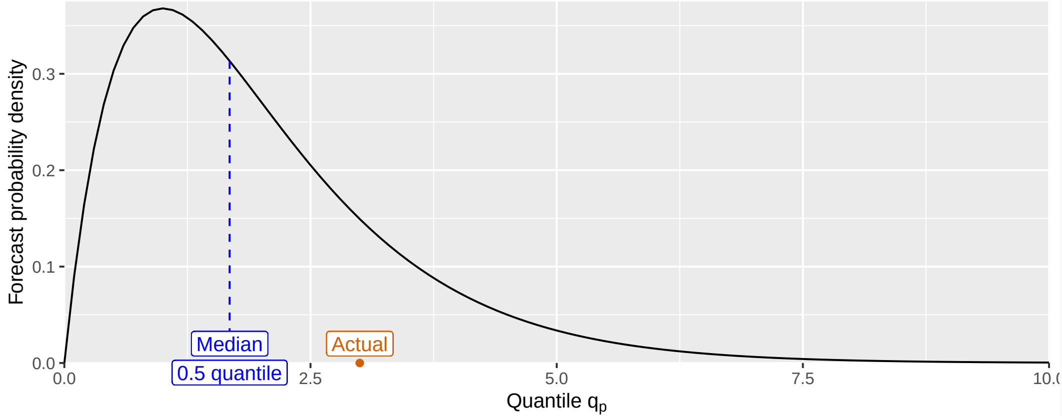
Visualising forecasts

```
tourism_fc %>%
  filter(Region == "Adelaide", Purpose=="Business", .model == "arima") %>%
  autoplot(tourism) +
  labs(title = "Adelaide Business Trips", y = "Thousands of visitors") +
  guides(color = guide_legend(title = "Forecast"))
```

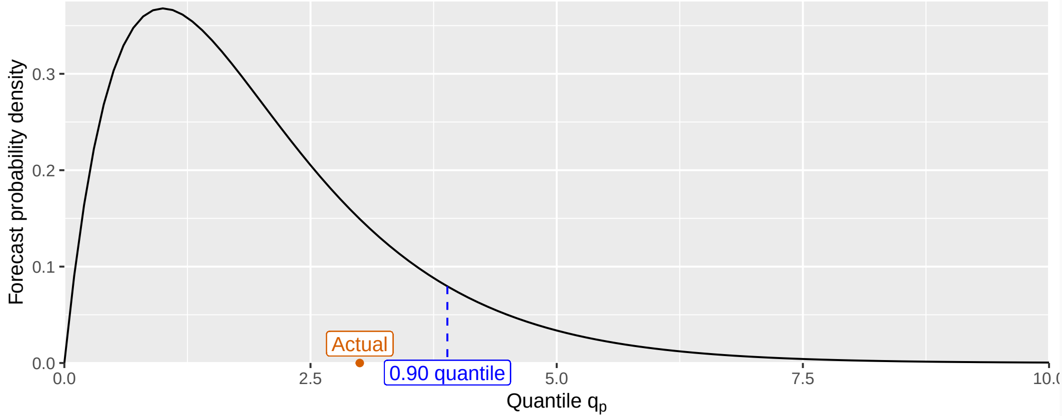


Outline

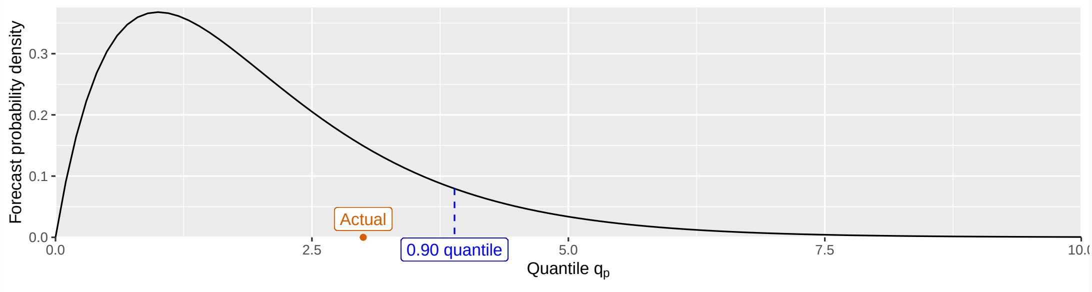
Evaluating probabilistic forecasts



Evaluating probabilistic forecasts



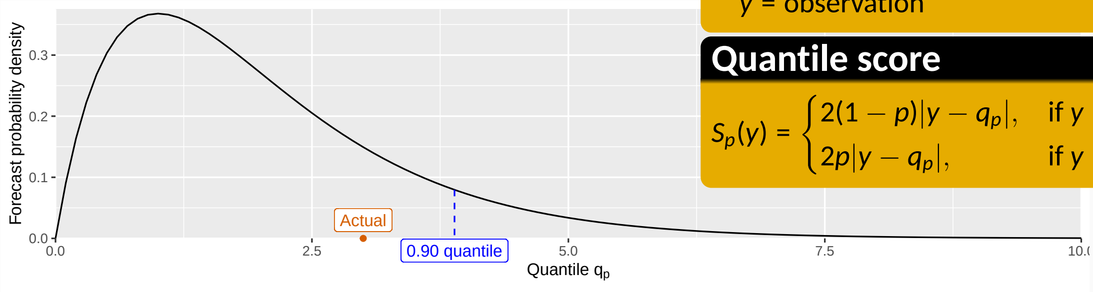
Evaluating probabilistic forecasts



Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

y = observation



Quantile score

$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \geq q_p \end{cases}$$

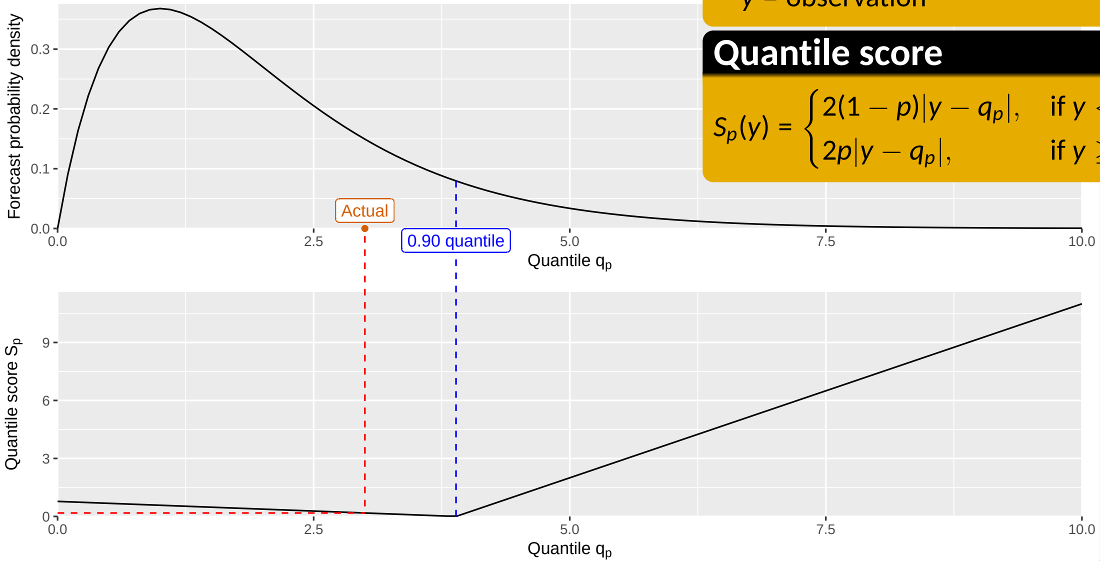
Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

y = observation

Quantile score

$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \geq q_p \end{cases}$$



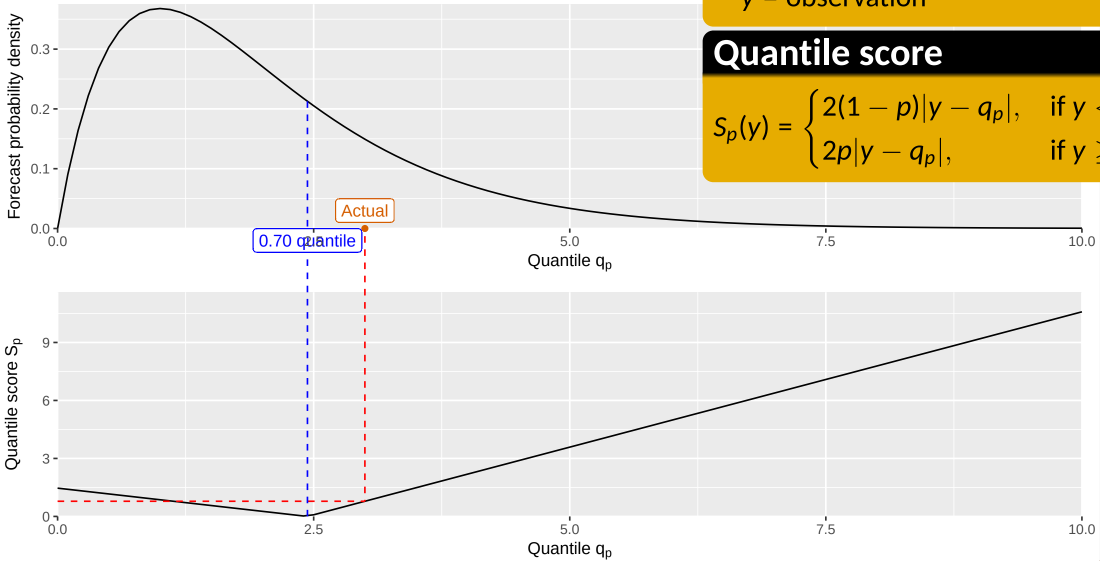
Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

y = observation

Quantile score

$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \ge q_p \end{cases}$$



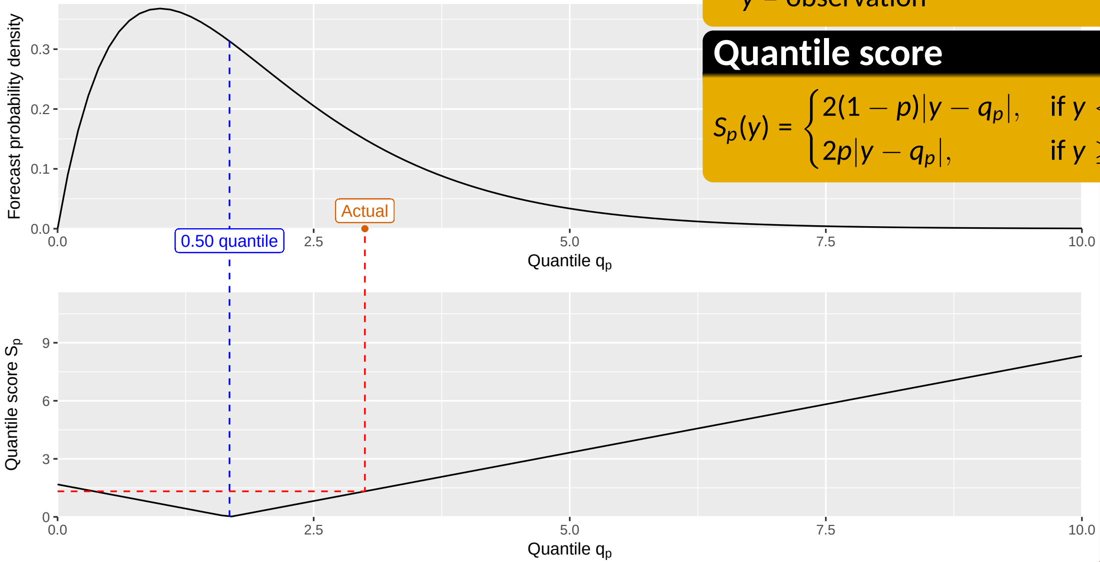
Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

y = observation

Quantile score

$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \geq q_p \end{cases}$$



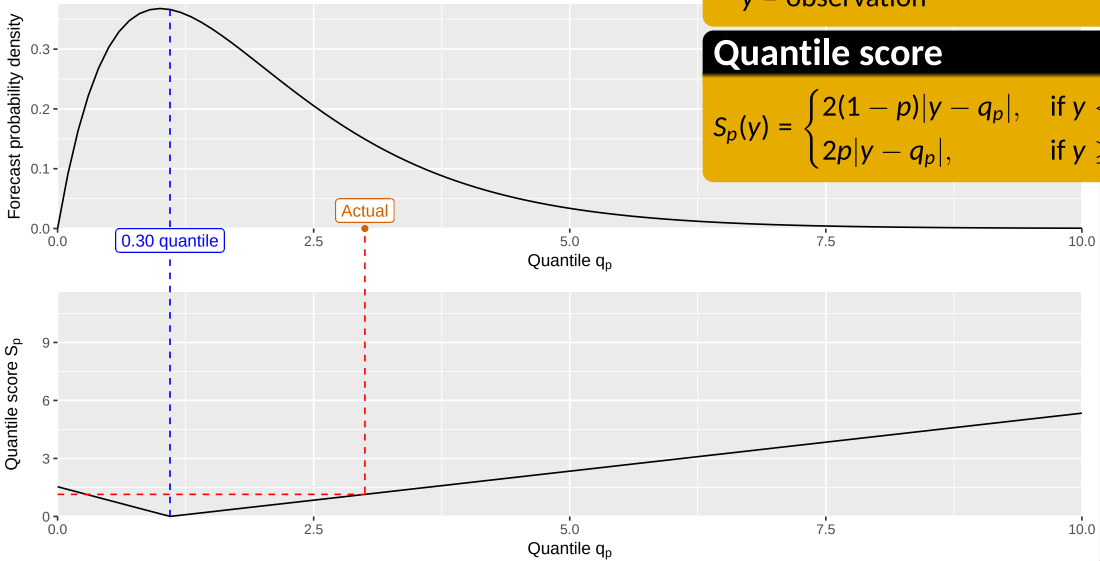
Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

y = observation

Quantile score

$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \geq q_p \end{cases}$$



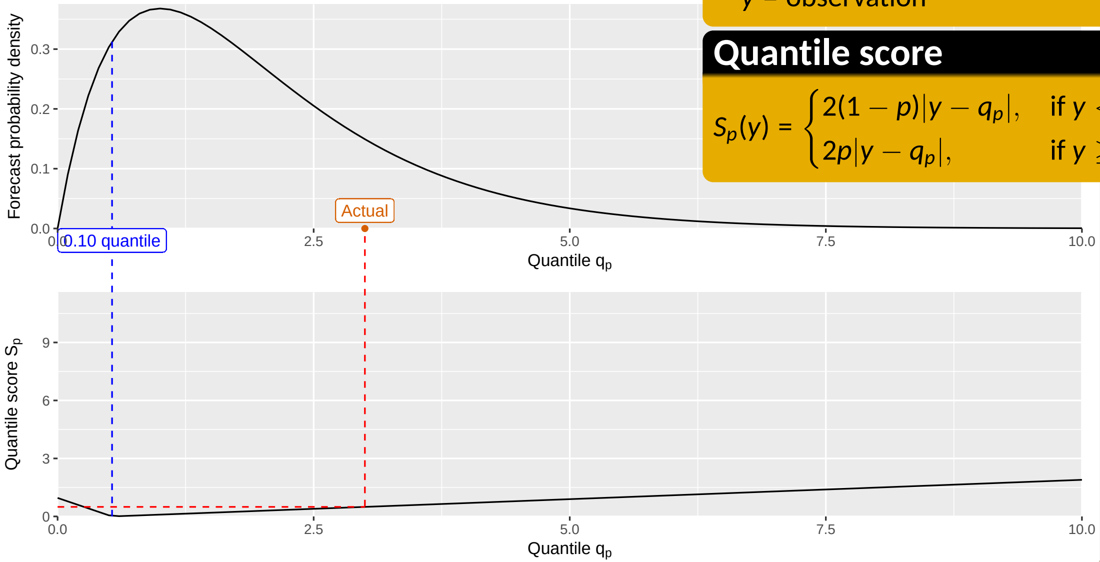
Evaluating probabilistic forecasts

q_p = quantile forecast with prob. p

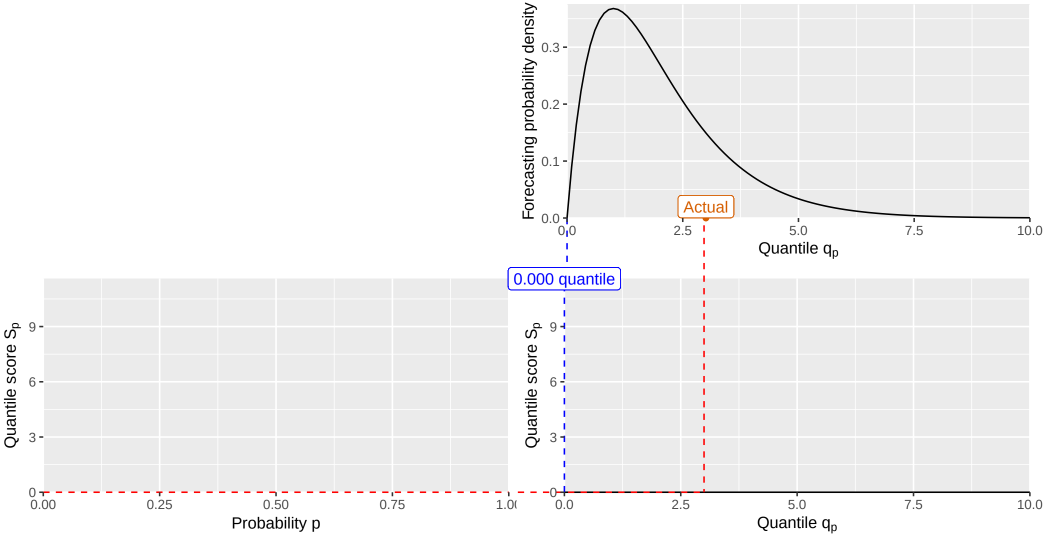
y = observation

Quantile score

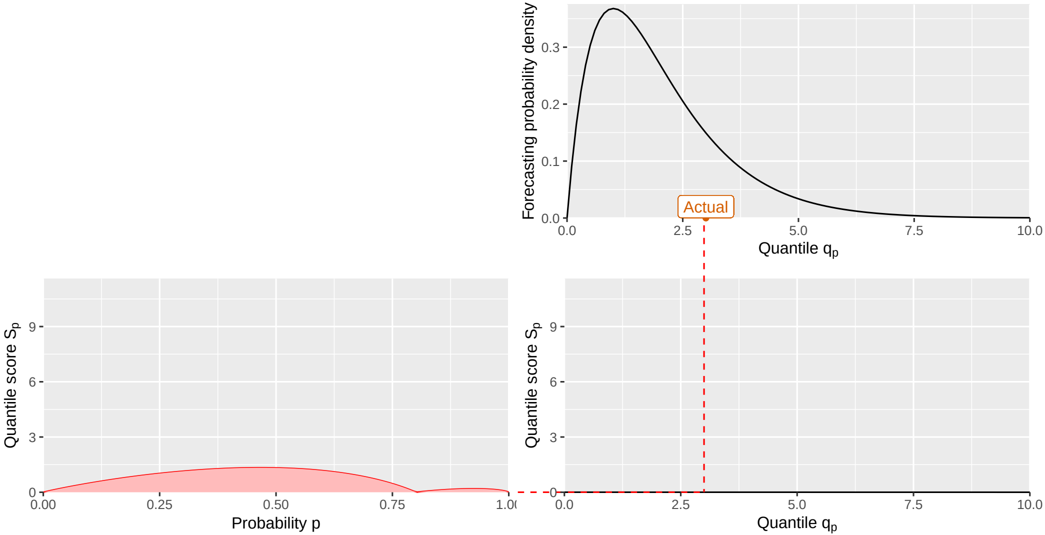
$$S_p(y) = \begin{cases} 2(1-p)|y-q_p|, & \text{if } y < q_p \\ 2p|y-q_p|, & \text{if } y \geq q_p \end{cases}$$



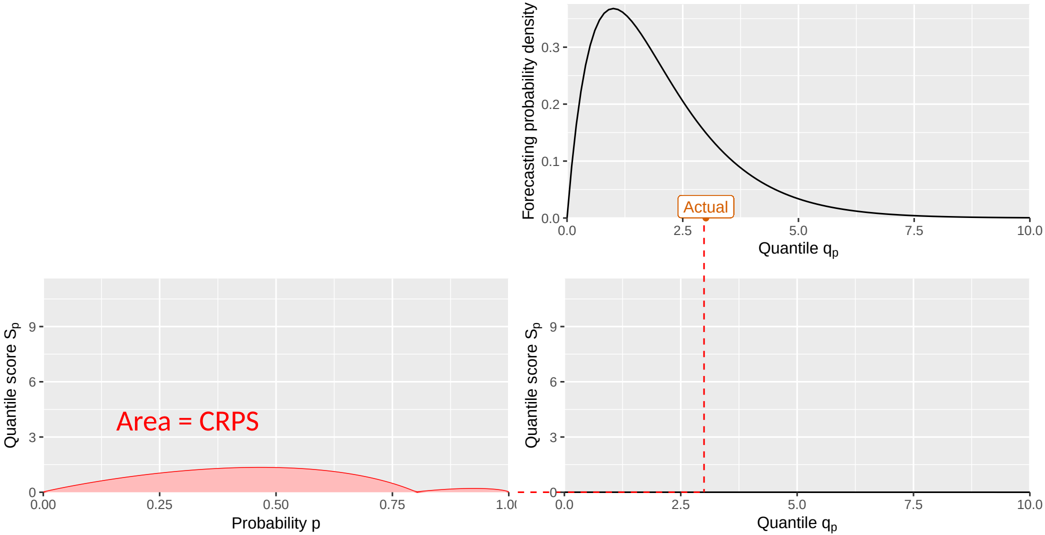
Evaluating probabilistic forecasts



Evaluating probabilistic forecasts



Evaluating probabilistic forecasts



Evaluating probabilistic forecasts

```r
tourism_fc %>%
  accuracy(tourism, measures = list(MSE=MSE, CRPS=CRPS))
```

```
## # A tibble: 912 x 7
##    .model Region         State Purpose  .type    MSE   CRPS
##    <chr>  <chr>          <chr> <chr>    <chr>  <dbl>  <dbl>
##  1 arima  Adelaide       SA    Business Test   840.   17.1
##  2 arima  Adelaide       SA    Holiday  Test   968.   18.1
##  3 arima  Adelaide       SA    Other    Test   188.    7.95
##  4 arima  Adelaide       SA    Visiting Test  1302.   21.4
##  5 arima  Adelaide Hills SA    Business Test    21.7   2.39
##  6 arima  Adelaide Hills SA    Holiday  Test    52.4   4.18
##  7 arima  Adelaide Hills SA    Other    Test     2.31  0.893
##  8 arima  Adelaide Hills SA    Visiting Test   144.    7.08
##  9 arima  Alice Springs  NT    Business Test   150.    7.78
## 10 arima  Alice Springs  NT    Holiday  Test    93.2   5.59
## # ... with 902 more rows
```

Evaluating probabilistic forecasts

```
tourism_fc %>%
  accuracy(tourism, measures = list(SS_MSE=skill_score(MSE), SS_CRPS=skill_score(CRPS)))
```

```
## # A tibble: 912 x 7
##    .model Region         State Purpose  .type SS_MSE  SS_CRPS
##    <chr>  <chr>          <chr> <chr>    <chr>  <dbl>    <dbl>
##  1 arima  Adelaide       SA    Business Test  -0.719 -0.00789
##  2 arima  Adelaide       SA    Holiday  Test   0.394  0.212
##  3 arima  Adelaide       SA    Other    Test   0.787  0.578
##  4 arima  Adelaide       SA    Visiting Test  -1.01  -0.354
##  5 arima  Adelaide Hills SA    Business Test   0.693  0.508
##  6 arima  Adelaide Hills SA    Holiday  Test  -0.568 -0.136
##  7 arima  Adelaide Hills SA    Other    Test   0.834  0.565
##  8 arima  Adelaide Hills SA    Visiting Test   0.120  0.0788
##  9 arima  Alice Springs  NT    Business Test  -1.43  -0.696
## 10 arima  Alice Springs  NT    Holiday  Test   0.468  0.186
## # ... with 902 more rows
```

Evaluating probabilistic forecasts

```
tourism_fc %>%
  accuracy(tourism, measures = list(SS_MSE=skill_score(MSE), SS_CRPS=skill_score(CRPS))) %>%
  group_by(.model) %>%
  summarise(SS_MSE = mean(SS_MSE), SS_CRPS=mean(SS_CRPS)) %>%
  arrange(desc(SS_CRPS))
```

```
## # A tibble: 3 x 3
##   .model   SS_MSE SS_CRPS
##   <chr>     <dbl>   <dbl>
## 1 ets      0.155   0.138
## 2 ensemble 0.141   0.138
## 3 arima    0.0636  0.0999
```

Tidyverts developers

Earo Wang

Mitchell O’Hara-Wild

More information

- Slides and papers: **robjhyndman.com**
- Packages: **tidyverts.org**
- Forecasting textbook using tidyverts package: **OTexts.com/fpp3**

Find me at ...

@robjhyndman

@robjhyndman

robjhyndman.com

rob.hyndman@monash.edu