Time Series Analysis & Forecasting in R

NY R Conference Workshop by Mitchell O'Hara

Recap by Alan Wu and Anjile An

Day 1 Recap

tidyverts packages

tsibble
Temporal data frames and tools

fable
Tidy forecasting

feasts
Feature extraction and statistics

tsibbledata
Diverse datasets for tsibble

fable.prophet
Interface to prophet forecaster for fable

tsibbletalk
Interactive crosstalk graphics for tsibble

tsibbles (time-series tibbles)

```
# A tsibble: 15,150 x 6 [1Y]
# Key:       Country [263]
    Year Country            GDP Imports Exports Population
   Index Key                Measured variables
 1  1960 Afghanistan  537777811.    7.02    4.13    8996351
 2  1961 Afghanistan  548888896.    8.10    4.45    9166764
 3  1962 Afghanistan  546666678.    9.35    4.88    9345868
 4  1963 Afghanistan  751111191.   16.9     9.17    9533954
 5  1964 Afghanistan  800000044.   18.1     8.89    9731361
 6  1965 Afghanistan 1006666638.   21.4    11.3     9938414
 7  1966 Afghanistan 1399999967.   18.6     8.57   10152331
 8  1967 Afghanistan 1673333418.   14.2     6.77   10372630
 9  1968 Afghanistan 1373333367.   15.2     8.90   10604346
10  1969 Afghanistan 1408888922.   15.0    10.1    10854428
# i 15,140 more rows
```

Other common time indexes:

- Annual
- Quarterly
- Monthly
- Weekly
- Daily
- Sub-daily

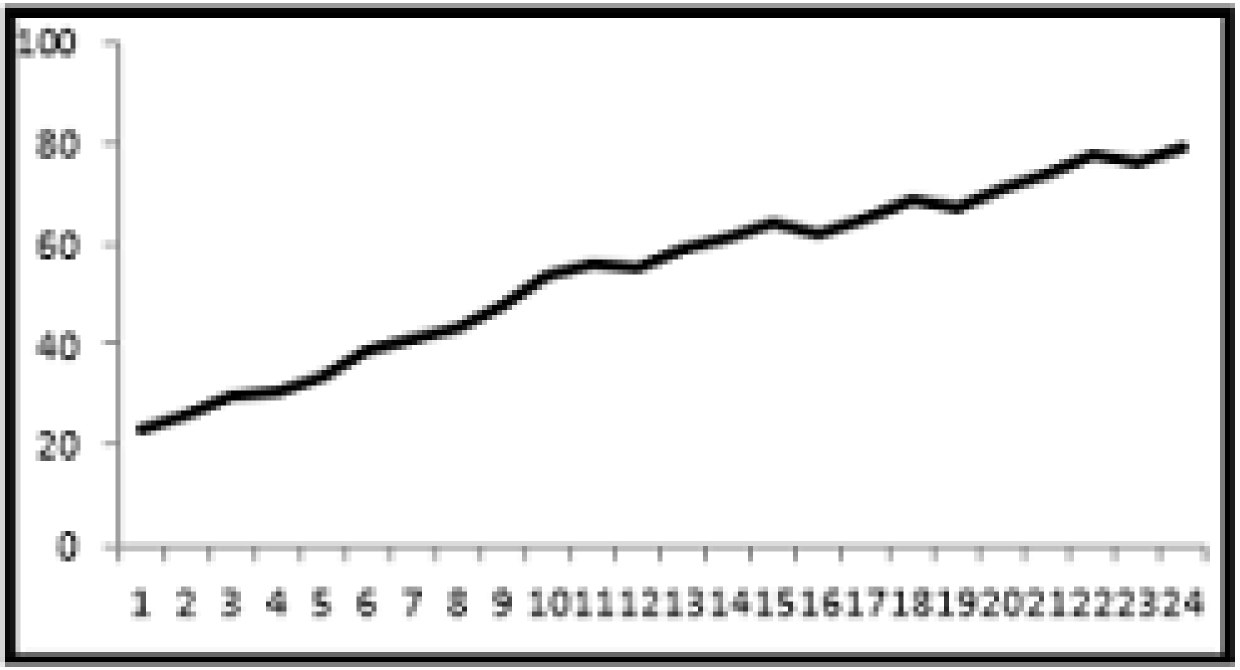
Loading a CSV into tsibble

```r
prison <- readr::read_csv("data/prison_population.csv") |>
  mutate(Quarter = yearquarter(date)) |>
  select(-date) |>
  as_tsibble(
    index = Quarter,
    key = c(state, gender, legal, indigenous)
  )
```

```
# A tsibble: 3,072 x 6 [1Q]
# Key:       state, gender, legal, indigenous [64]
   state gender legal    indigenous count Quarter
   <chr> <chr>  <chr>    <chr>      <dbl>   <qtr>
 1 ACT   Female Remanded ATSI           0 2005 Q1
 2 ACT   Female Remanded ATSI           1 2005 Q2
 3 ACT   Female Remanded ATSI           0 2005 Q3
 4 ACT   Female Remanded ATSI           0 2005 Q4
 5 ACT   Female Remanded ATSI           1 2006 Q1
 6 ACT   Female Remanded ATSI           1 2006 Q2
```

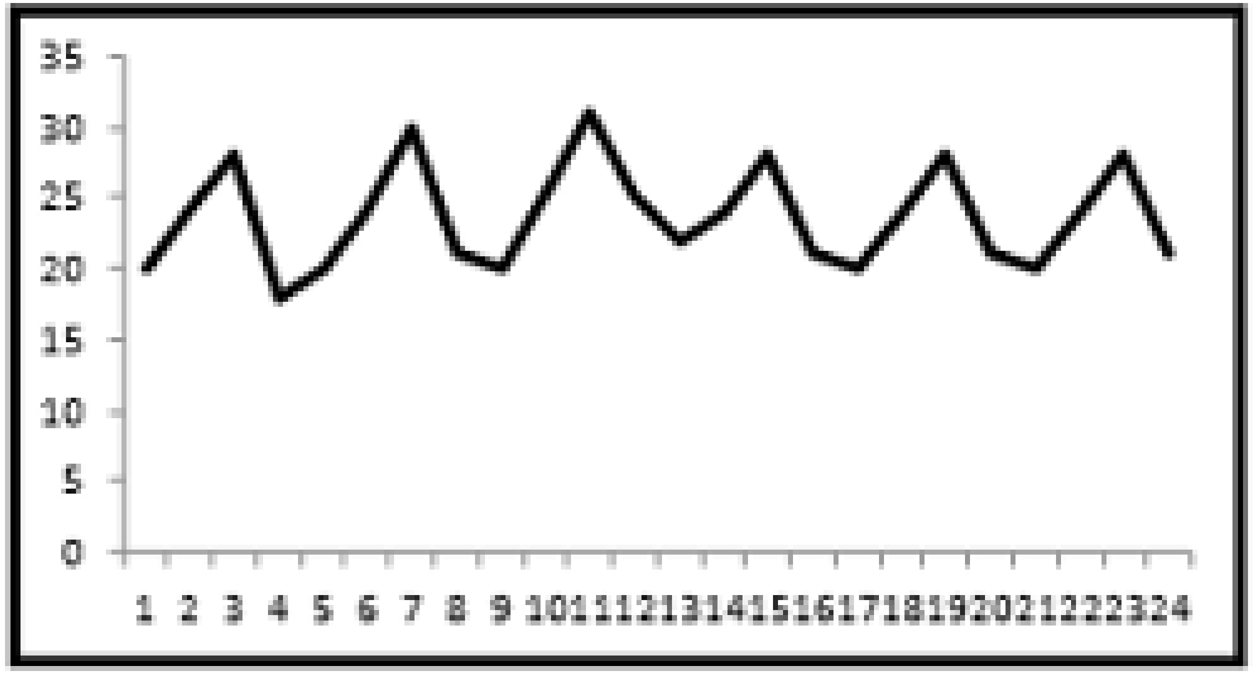
Time series patterns

- **Trend:** pattern exists when there is a long-term increase or decrease in the data.



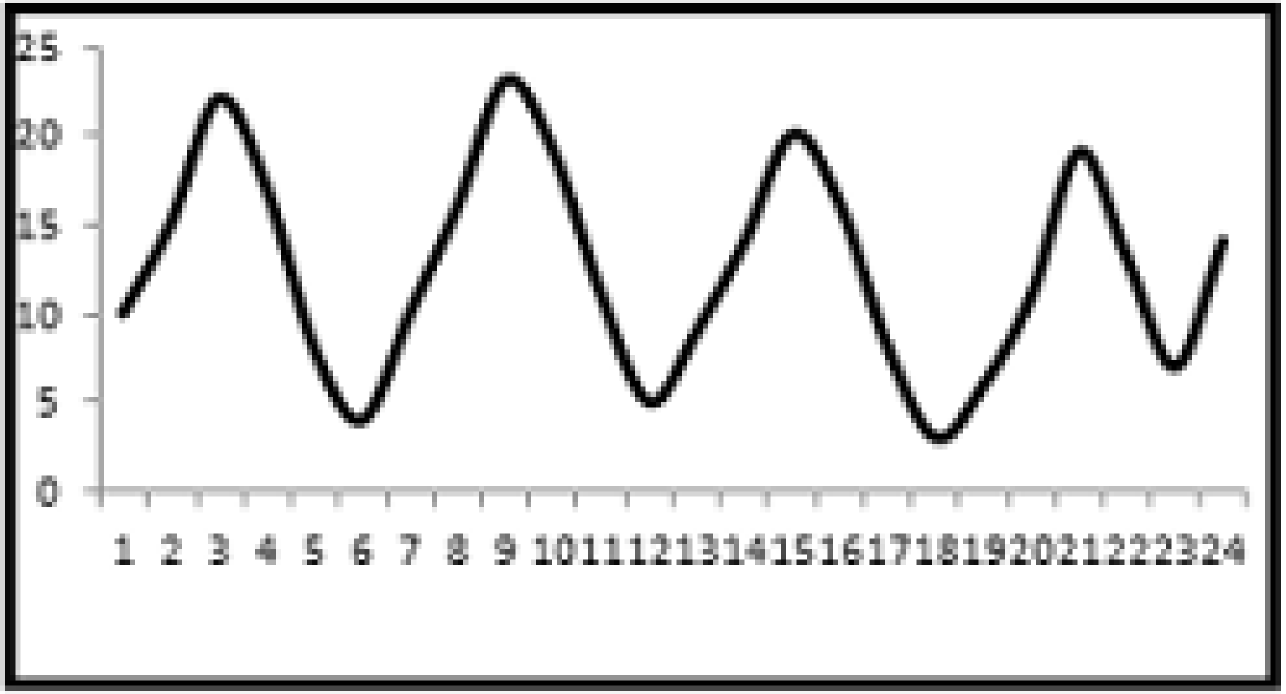
Time series patterns

- **Seasonal:** pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).



Time series patterns

- **Cyclic:** pattern exists when data exhibit rises and falls that are not of fixed period (duration usually of at least 2 years).



Seasonal vs cyclic

Differences between seasonal and cyclic patterns:

1. Seasonal pattern constant length; cyclic pattern variable length
2. Average length of cycle longer than length of seasonal pattern
3. Magnitude of cycle more variable than magnitude of seasonal pattern

Seasonal vs cyclic

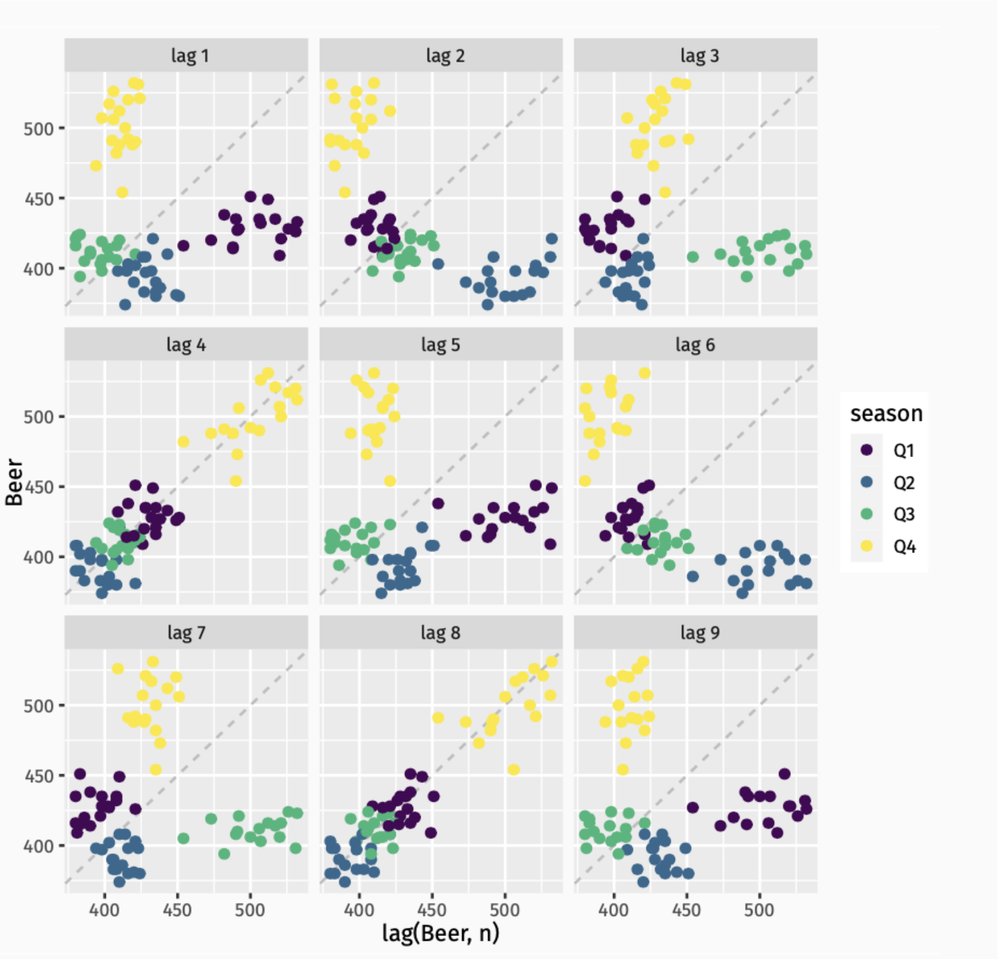
Differences between seasonal and cyclic patterns:

1. Seasonal pattern constant length; cyclic pattern variable length
2. Average length of cycle longer than length of seasonal pattern
3. Magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Lagged plots and autocorrelation

- Each graph shows y_t plotted against y_{t-k} for different values of *k*.
- The autocorrelations (AutoCorrelation Function) are the correlations associated with these scatterplots
  - r_1 = Correlation(y_t, y_{t-1})
  - r_2 = Correlation(y_t, y_{t-2})
  - etc.

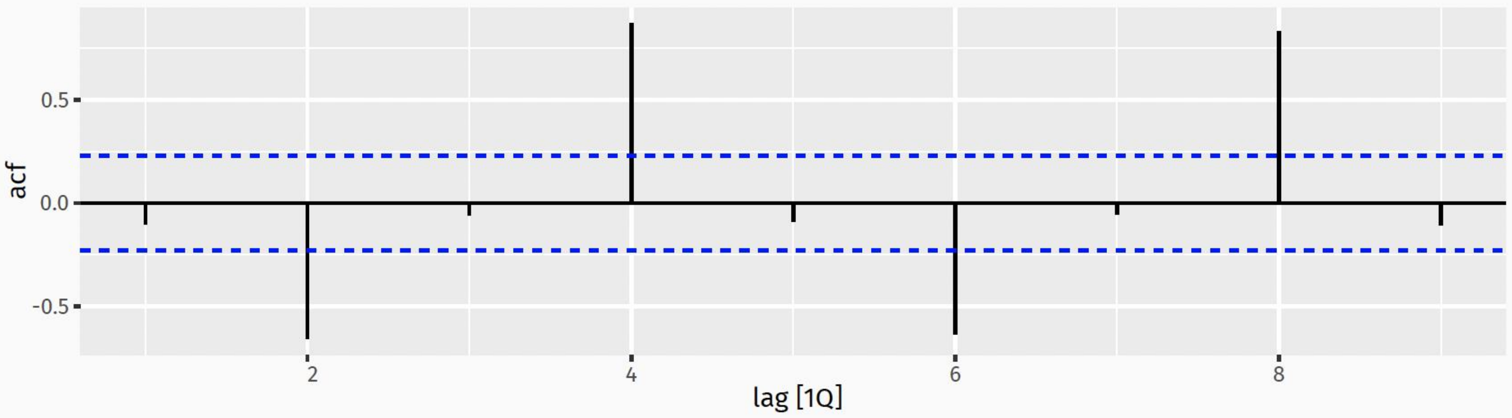


Lagged scatterplots

If there is seasonality, the ACF (autocorrelation function) at the seasonal lag (e.g., 12 for monthly data) will be large and positive.

Results for first 9 lags for beer data:

```
new_production |>
  ACF(Beer, lag_max = 9) |>
  autoplot()
```



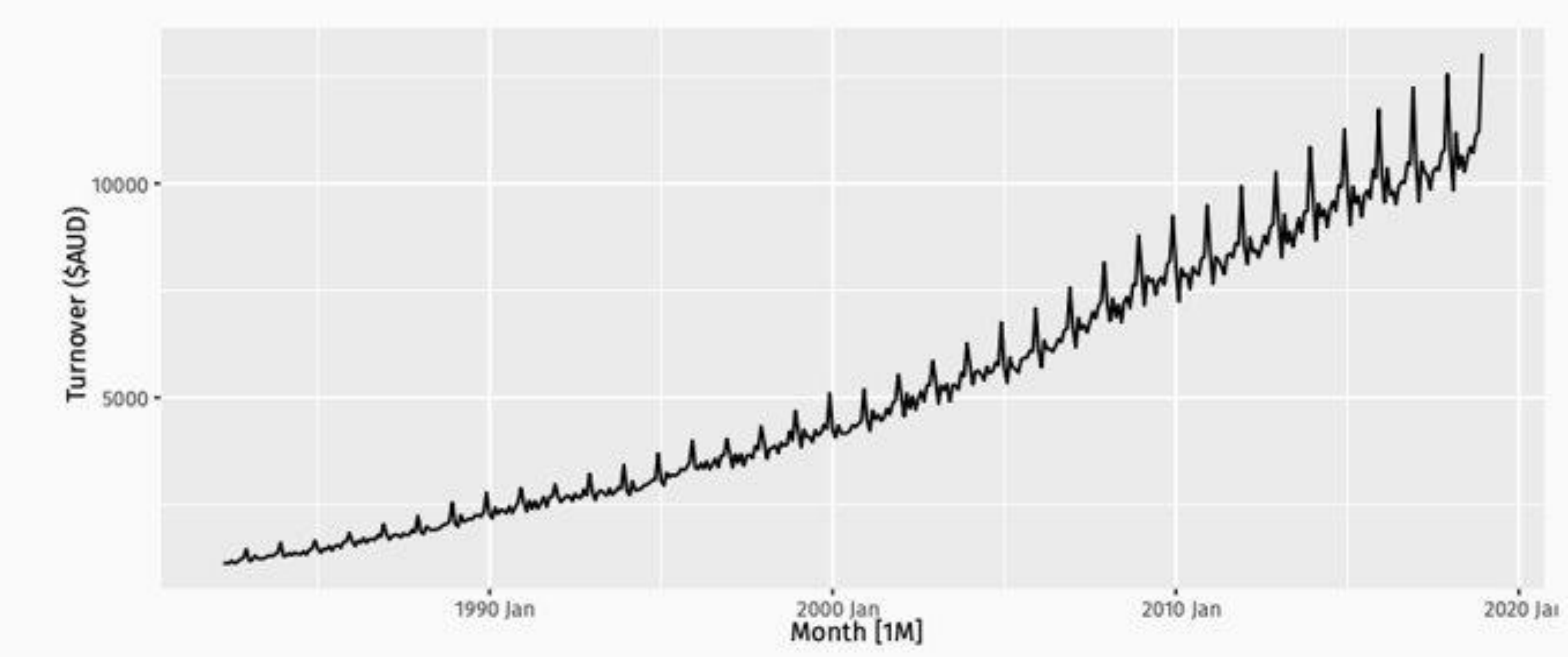
Day 2 Recap

Transformations

- Transformations can be useful if different variations exist in the time series

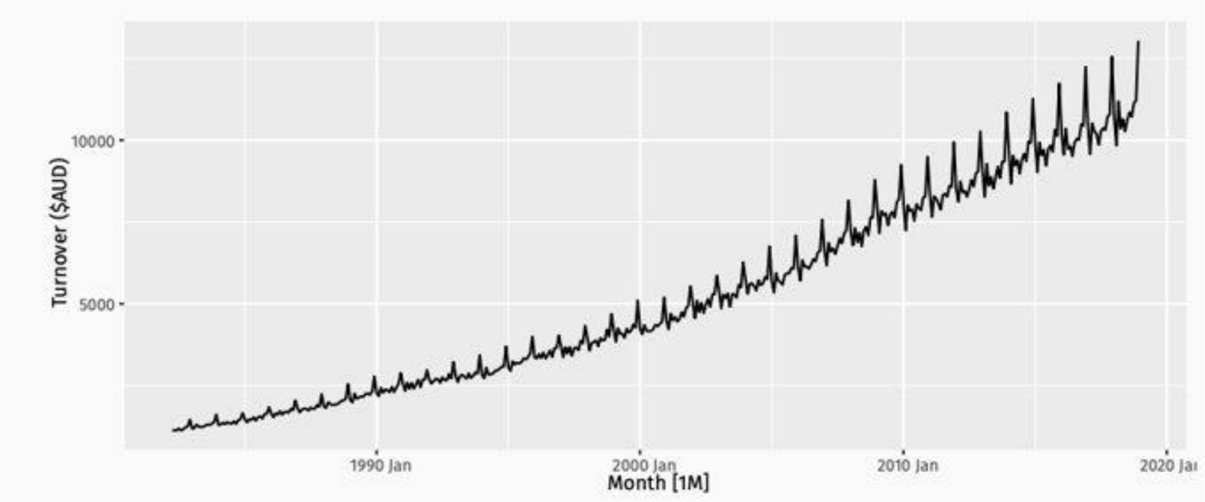
Transformations

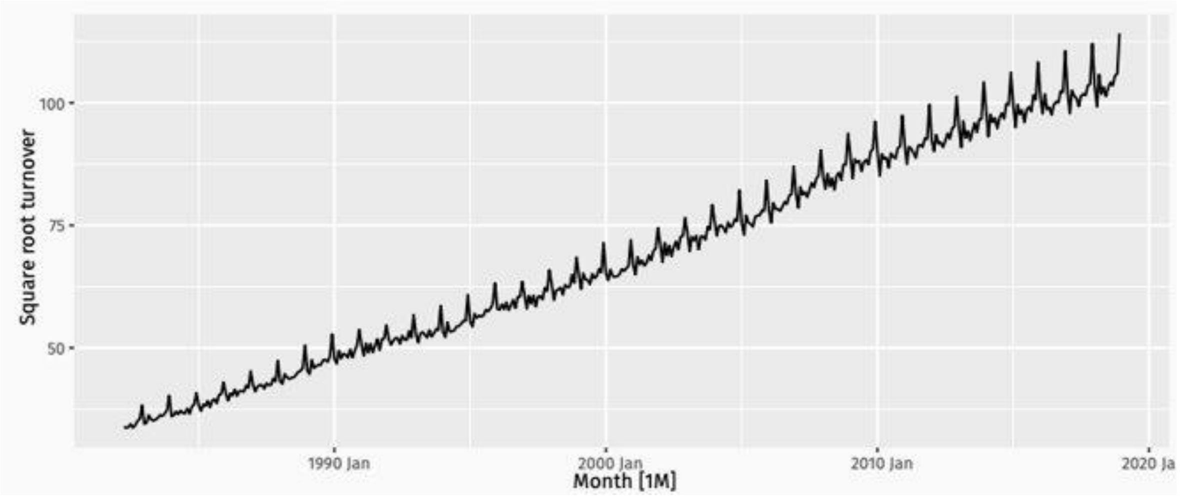
- Transformations can be useful if different variations exist in the time series



Transformations

- Transformations can be useful if different variations exist in the time series

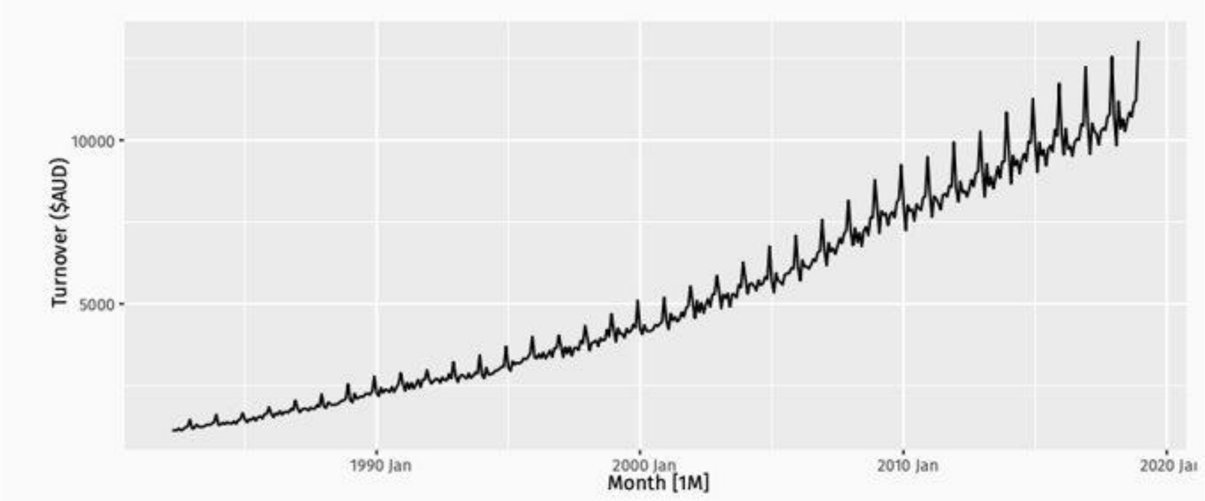


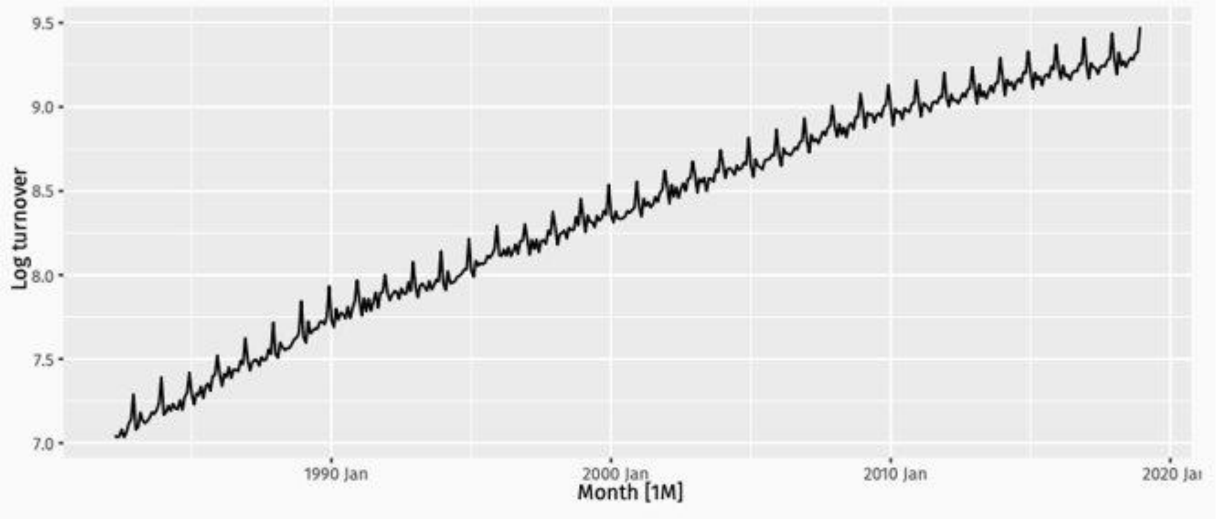


Square root

Transformations

- Transformations can be useful if different variations exist in the time series





Cube root

Transformations

- In general, Box-Cox transformations allow us to produce equal variance without guess and check

```
food |>
  features(Turnover, features = guerrero)

# A tibble: 1 x 1
  lambda_guerrero
            <dbl>
1          0.0895
```

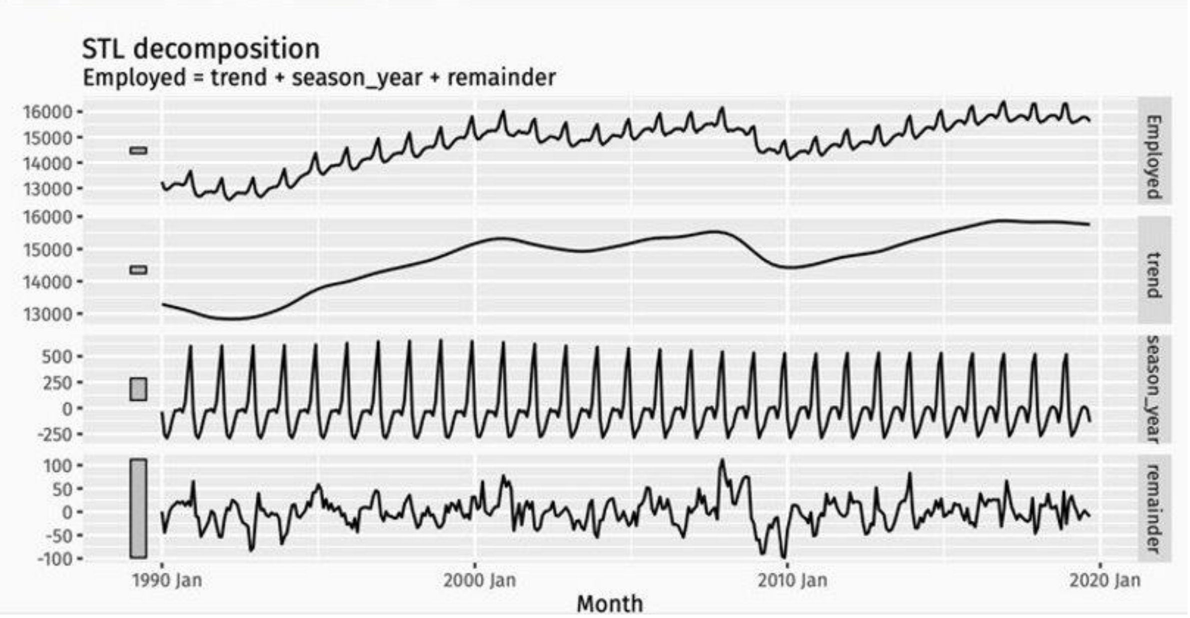
Decomposition

- Trends can be decomposed into seasonal, trend, and random components
  - Trend-cycle: aperiodic changes in level over time
  - Seasonal: periodic changes due to seasonal factors
  - Remainder: noise

Decomposition

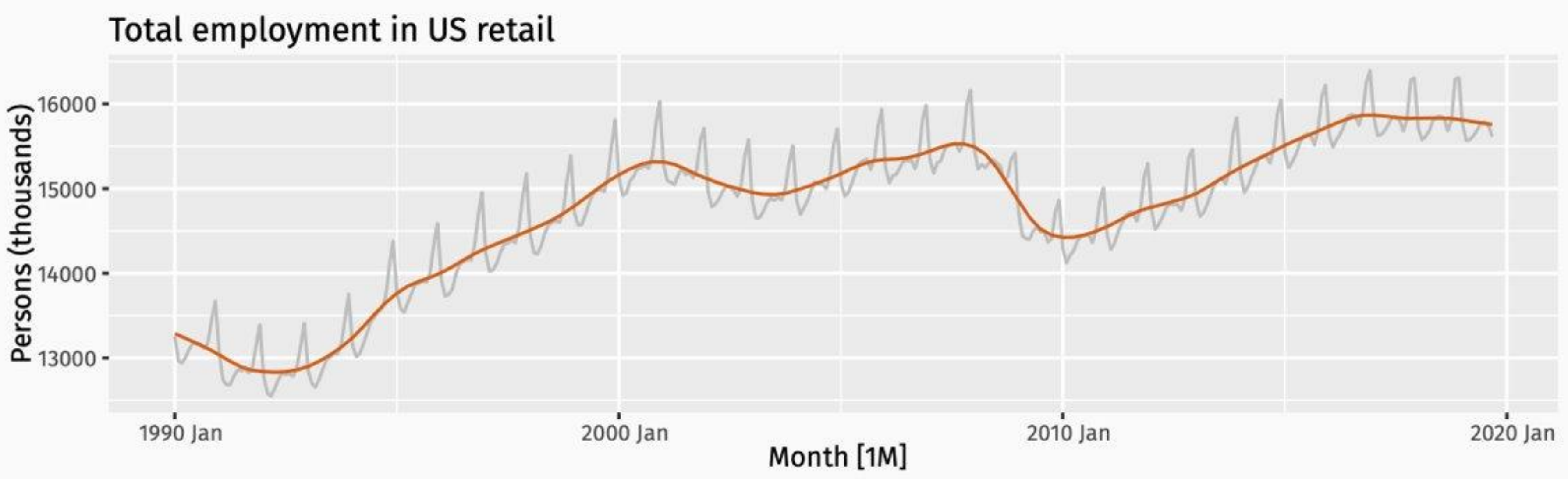
- Trends can be decomposed into seasonal, trend, and random components
  - Trend-cycle: aperiodic changes in level over time
  - Seasonal: periodic changes due to seasonal factors
  - Remainder: noise

```
components(dcmp) |> autoplot()
```



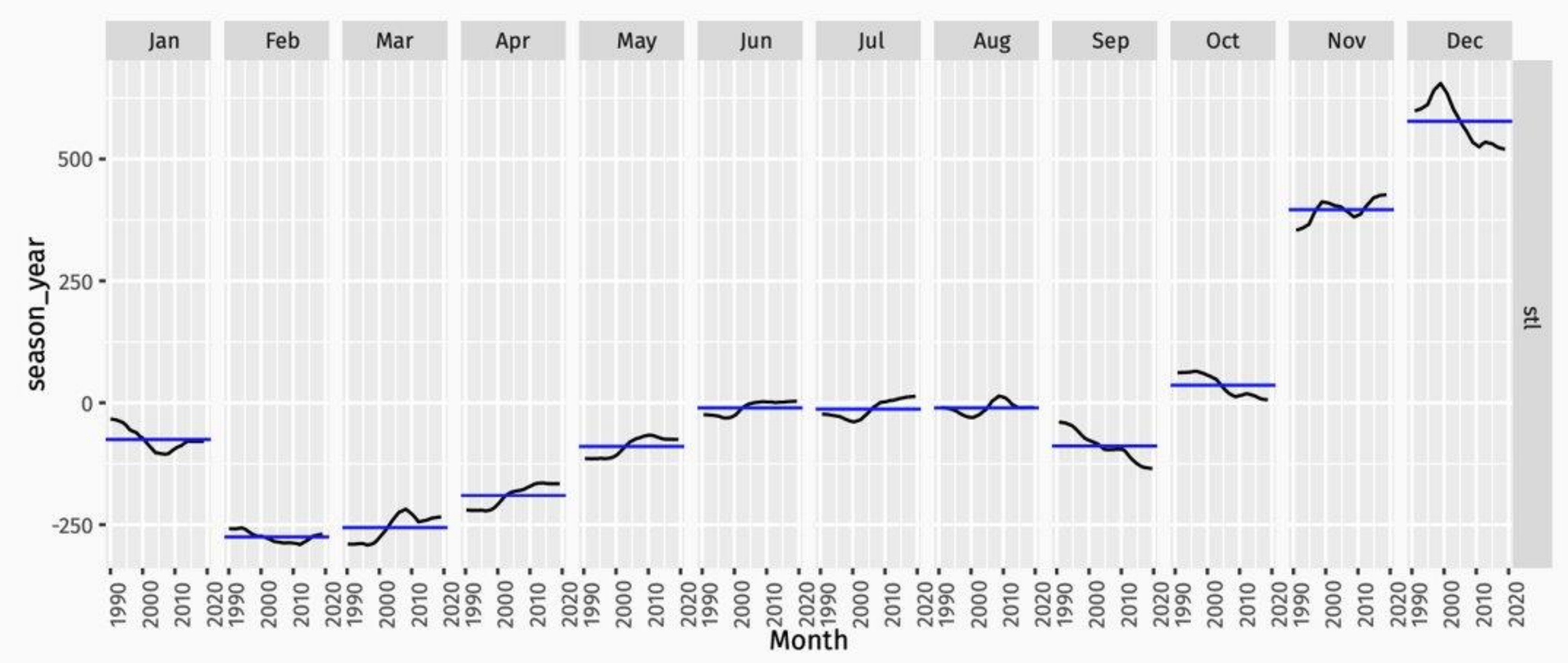
Decomposition

```
us_retail_employment |>
  autoplot(Employed, color = "gray") +
  autolayer(components(dcmp), trend, color = "#D55E00") +
  labs(y = "Persons (thousands)", title = "Total employment in US retail")
```



Decomposition

```r
components(dcmp) |> gg_subseries(season_year)
```



Further reading

https://otexts.com/fpp3/

Forecasting: Principles and Practice (3rd ed)

Rob J Hyndman and George Athanasopoulos

Monash University, Australia