Timeseries analysis with R

JIGSAW ACADEMY

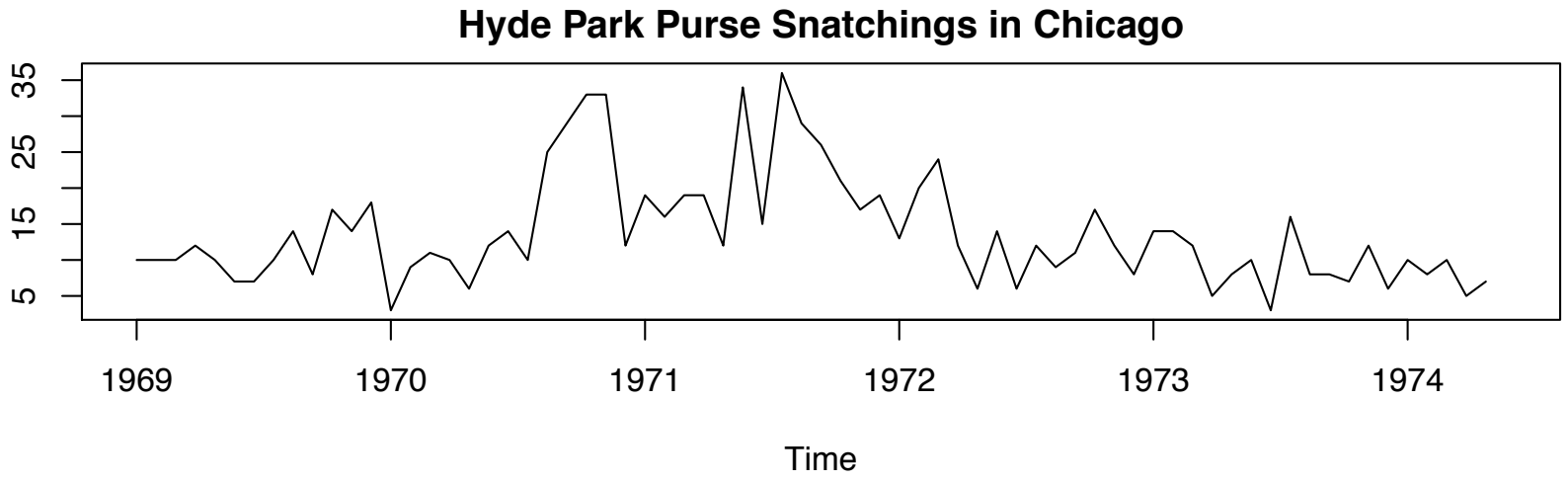
Analytics for Professionals

Outline

1. What is a timeseries?
2. Timeseries characteristics
3. Approaches to modeling timeseries
4. Automation in timeseries modeling

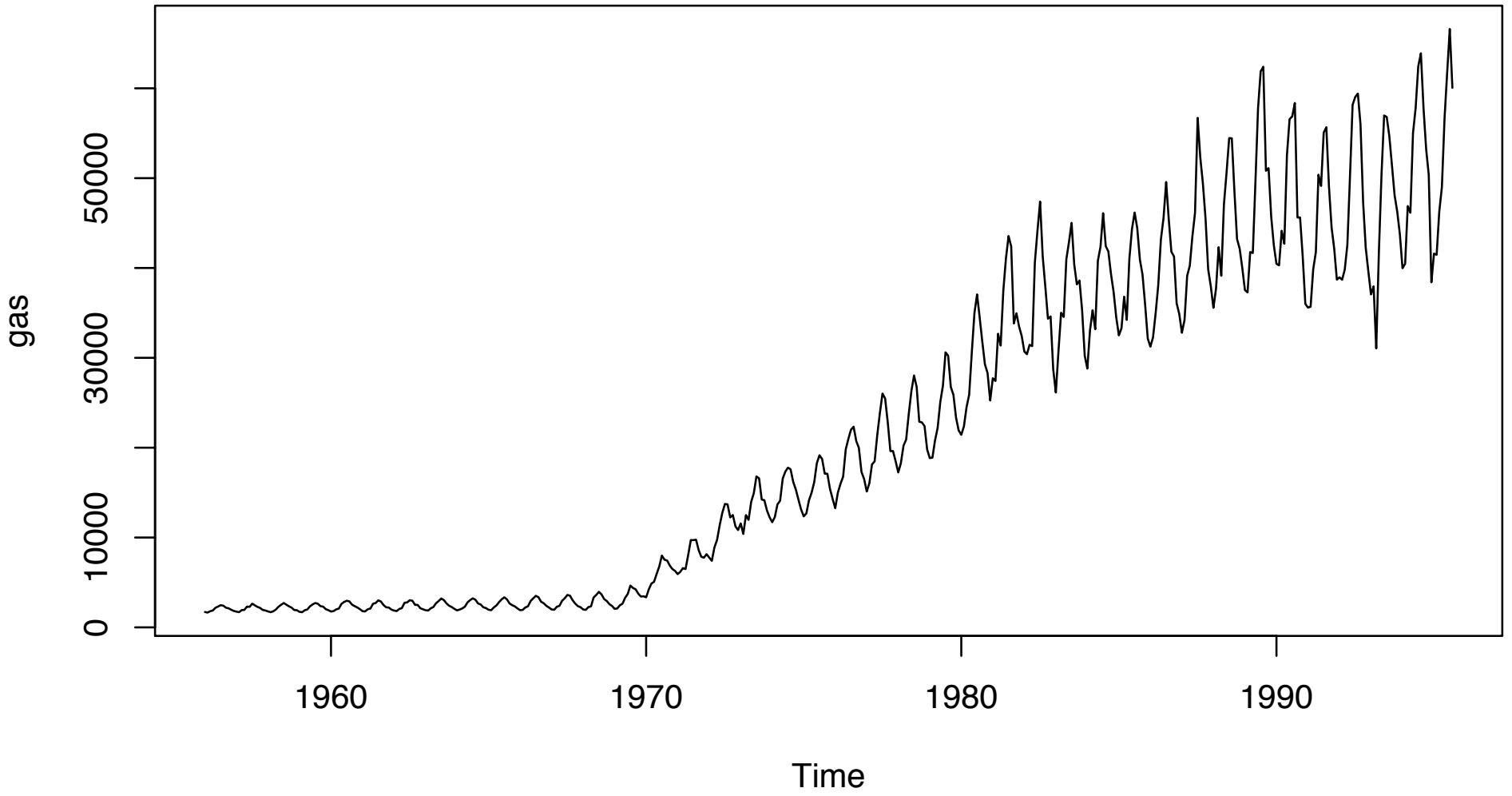
What is a timeseries?

- A phenomenon observed over time
- Observations on the same unit, system etc over time
- Therefore the observations are not independent
- Regular or irregular time intervals
  - temperature at 2 pm
  - stock price at 2 pm is actually last traded price, which itself is a random event

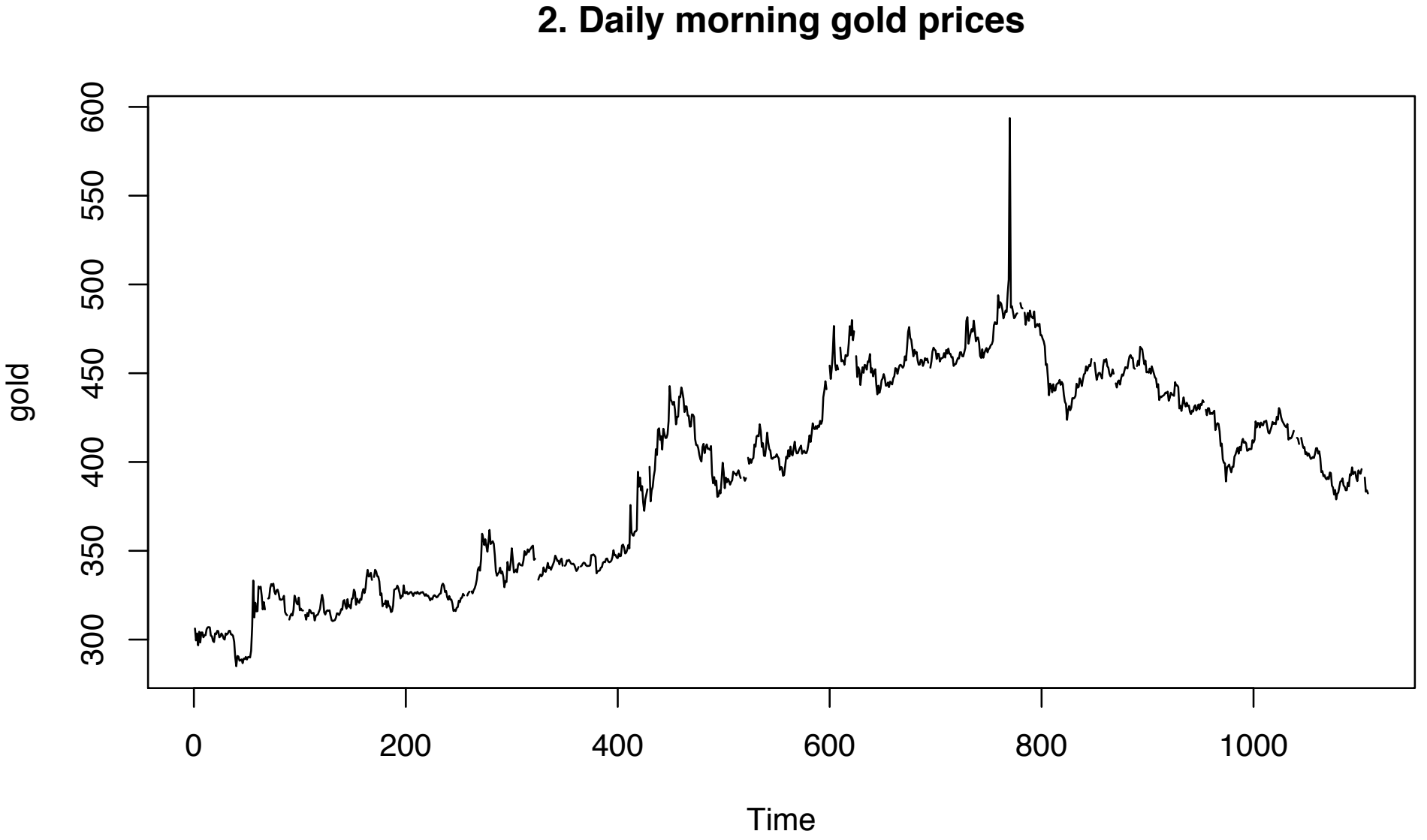


Timeseries characteristics

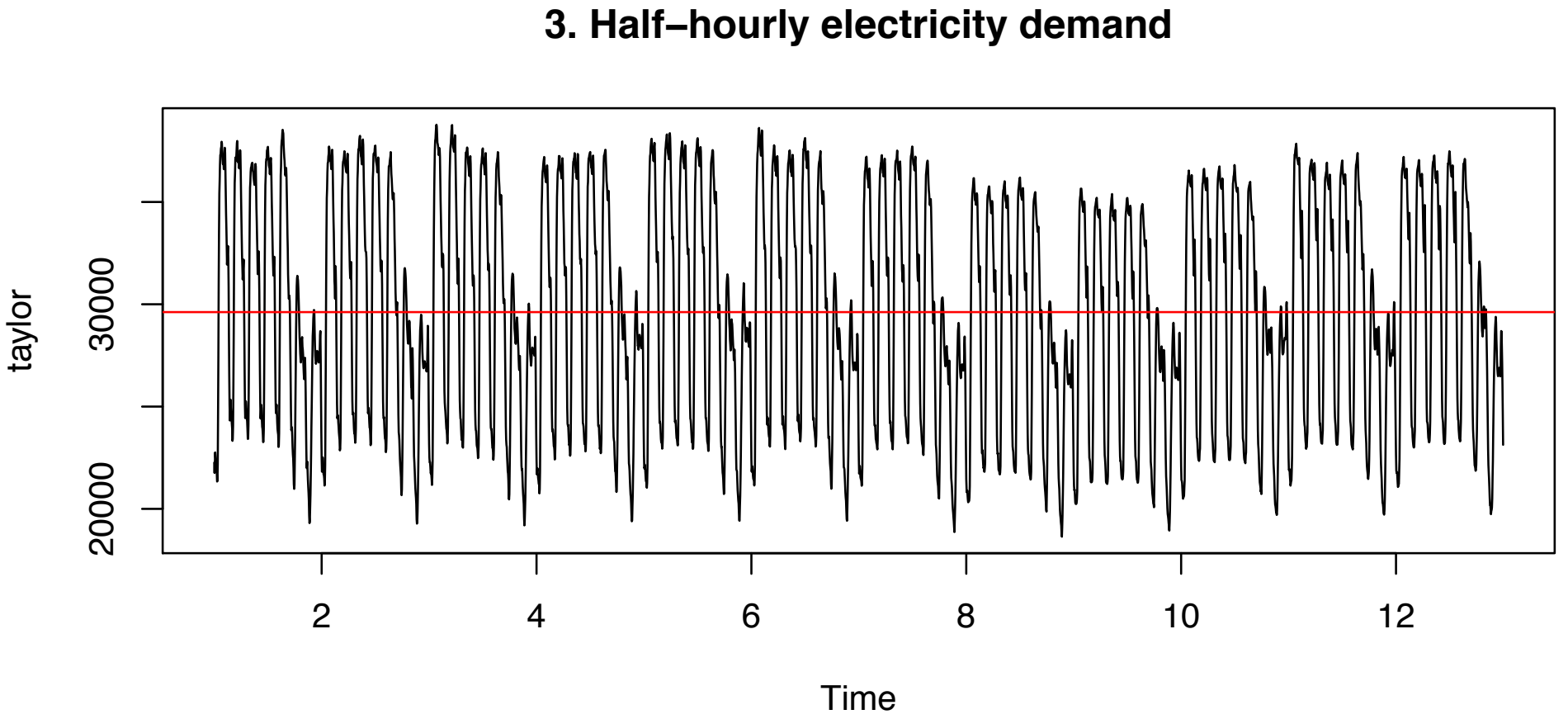
1. Australian monthly gas production



Timeseries characteristics



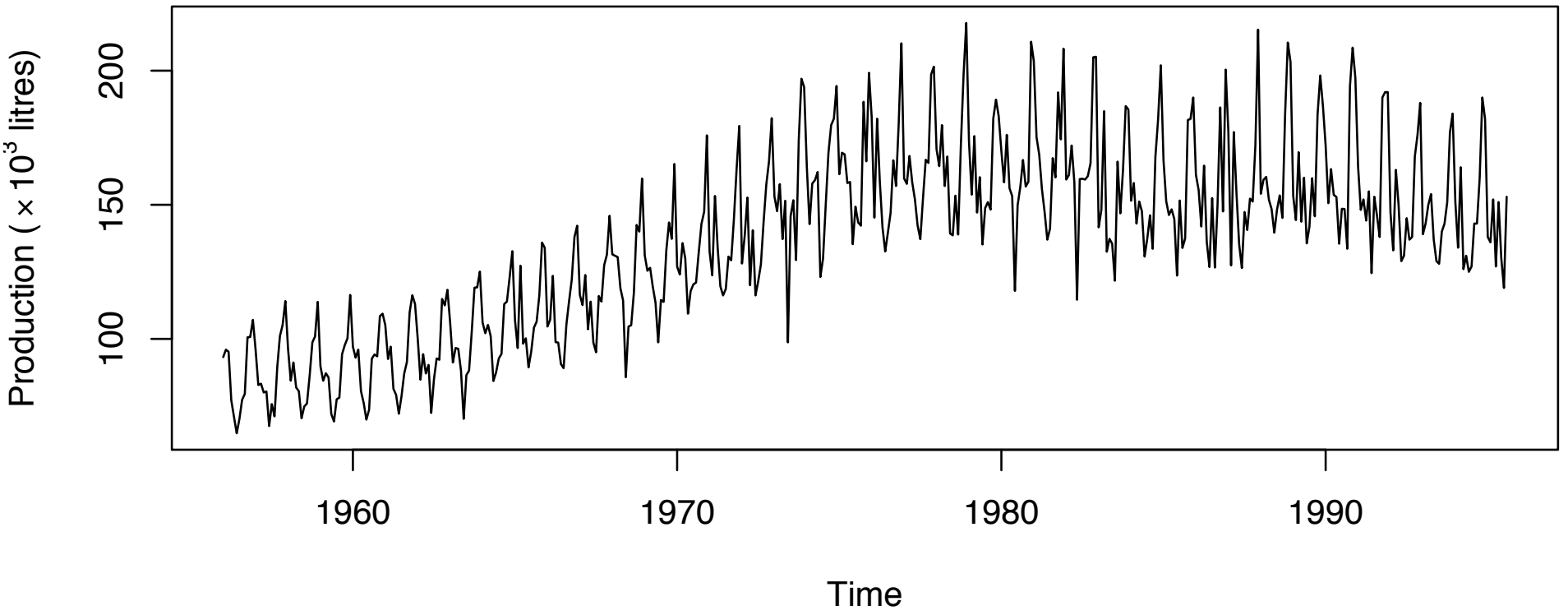
Timeseries characteristics



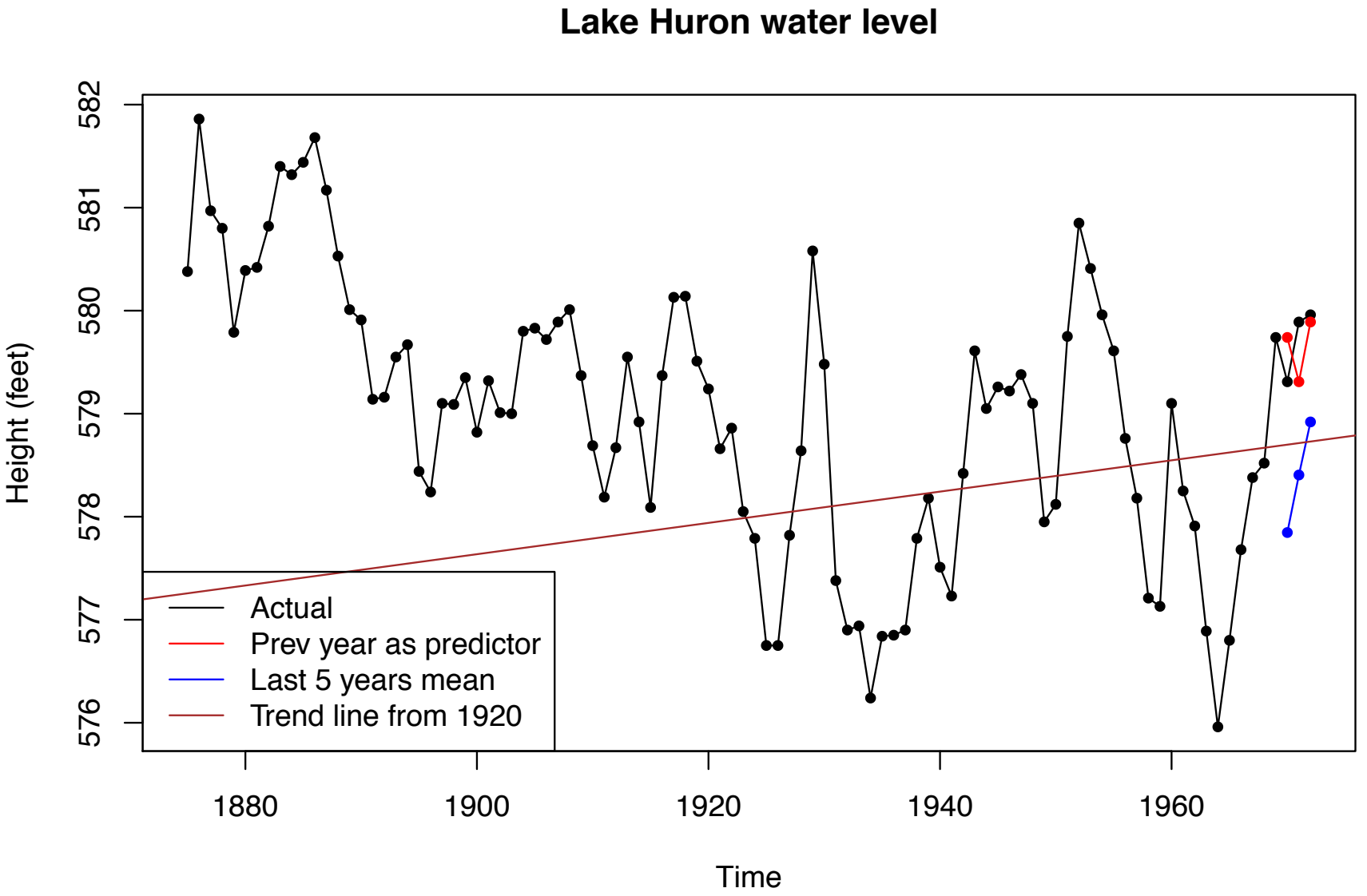
- Trend
- Seasonality

Timeseries object in R

- Beer production in Australia from Jan. 1956 to Aug. 1995
- http://134.76.173.220/beer.zip



Naive predictors



Exponential smoothing

- Average of previous values, with the most recent values weighted more strongly.

$$y_t = \alpha(y_{t-1}) + \alpha(1-\alpha)y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \cdots, \quad 0 < \alpha < 1$$

- Exponentially Weighted Moving Average (EWMA)
- `HoltWinters(x, beta=FALSE, gamma=FALSE)`

HoltWinters

HoltWinters offers following ways of modeling trend and seasonality

- Trend: None or Additive
- Seasonality: None, Additive or Multiplicative
- Any combinition of these is possible
- Only non-negative terms may be modelled by multiplicative seasonality
- Error term is always additive

ExponenTial Smoothing

ETS family of models offer the following

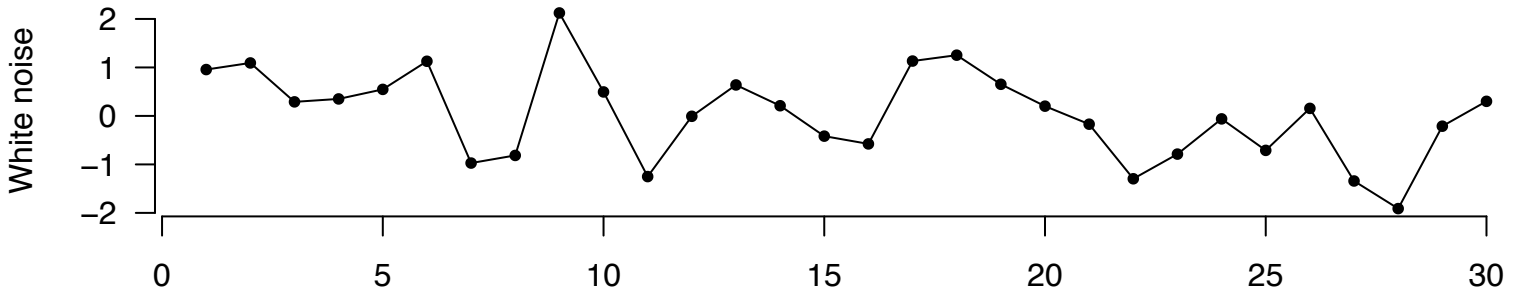
- Trend: None, Additive, Additive-damped, Multiplicative, Multiplicative-damped
- Seasonality: None, Additive or Multiplicative
- Error: Additive or Multiplicative
- In all 24 models are possible
- Identified using 3-letter combination like (M Ad M)

```
ets(y, model="ZZZ", damped=NULL)
```

Stationary timeseries

A timeseries (or a process as it is also called) is stationary if

1. The mean is constant over time
2. The variance is constant over time
3. Covariance between y_t and y_{t-h}, where h is the lag, depends only on h, and not on t, the time at which the covariance is measured.



- Better known as 'Covariance stationary'
- Augmented Dickey-Fuller test for stationarity: `tseries::adf.test`

Transformations for Stationarity

- Box-Cox transformation

$$y_t^{(\lambda)} = \begin{cases} \log(y_t), & \lambda = 0 \\ \frac{y_t^{\lambda}-1}{\lambda}, & \lambda \neq 0 \end{cases}$$

`MASS::boxcox`

- Differencing

$$\nabla y_t = y_t - y_{t-1} \ldots \text{first difference}$$

Box-Jenkins (ARIMA) Models

- Autoregressive process: Next term is a sum of some past terms

$$y_t = \eta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon$$

  - Stationarity requirement: $\sum \phi_i < 1$

- Moving average process: Next term is a sum of random shocks over time

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

  - Invertibility requirement: MA process $\sim$ AR process

- Real process is a combination of AR & MA processes
- order is the number of terms in the model
- ARIMA: Auto Regressive Integrated Moving Average

ARIMA modeling

- stats:: ar, arma, arima
- forecast:: Arima

```
Arima(x, order=c(0,0,0),
      seasonal=list(order=c(0,0,0), period=NA)
```

- forecast:: Acf, Pacf

ARIMA model identification

Parameters	ACF	PACF
AR(1)	exponential decay	spike at lag 1, no correlation for other lags
AR(2)	sine-wave shape pattern or a set of exponential decays	spikes at lags 1 and 2; no correlation for other lags
MA(1)	spike at lag 1, no correlation for other lags	damps out exponentially
MA(2)	spikes at lags 1 and 2, no correlation for other lags	sine-wave shape pattern or a set of exponential decays
AR(1), MA(1)	exponential decay starting at lag 1	exponential decay starting at lag 1

Automated Modeling

- Large number of models, repeatedly, every month or so
- Difficulty in choosing model
- Limited skill for bulk work
- `ets(y, model="ZZZ")`
- `auto.arima(y)`

Advanced Topics

- Packages: zoo, xts
  - irregular timeseries
  - multivariate & financial
- quantmod
  - financial timeseries
  - fetching series
  - custom plots
  - Re-purposed techniques & models

Cran: TaskViews

http://cran.r-project.org/web/views/TimeSeries.html