Data Preprocessing

# Nagel-Datensatz

## Pakete laden

## Oberflächliche Betrachtung

### Einladen

### Dim, classes, missings, range, mean + var

### Was tun mit missings?

## Plots

### Gewicht über Zeit

#### Wochenweise

#### Jahresweise

#### Ohne Feiertage?

### Volumen?

### Die unabhängigen Variablen?

# Umwandeln in cal

## Zunächst Variablenklassen korrigieren

## Quartale korrigieren

# Exogene Variablen anfügen

### Schulferien

### GDP

### Benzinpreis

# Stationarität gewährleisten

## Log

## Differencing

### *Step 2*:  Do any necessary differencing. The general guidelines are:

### If there is seasonality and no trend, then take a difference of lag *S*. For instance, take a 12th difference for monthly data with seasonality.  Seasonality will appear in the ACF by tapering slowly at multiples of S.

### If there is linear trend and no obvious seasonality, then take a first difference.  If there is a curved trend, consider a transformation of the data before differencing.

### If there is both trend and seasonality, apply a seasonal difference to the data and then re-evaluate the trend.  If a trend remains, then take first differences.  For instance, if the series is called *x*, the commands in R would be:

### diff12=diff(x, 12)

### plot(diff12)

### acf2(diff12)

### diff1and12 = diff(diff12, 1)

### If there is neither obvious trend nor seasonality, don’t take any differences.

## Train-Validation-Test-Split!