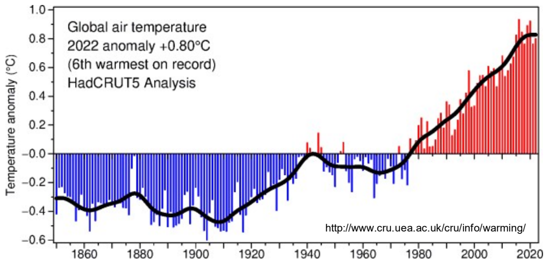
Global change research methods

II. Variability in time

Intro to the time series analysis

Using time series, we can examine the **dynamics** of phenomena over time.



Knowing the time series dynamics is fundamental for:

a) revealing the causes that are behind the time series development

b) reconstructing their behavior in the past

c) predicting their future development.

Time series compilation

Primary (data directly measured) and secondary (derived) sources

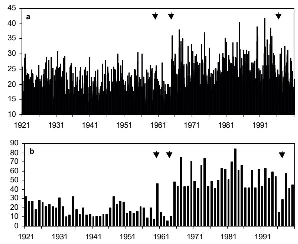
Possible problems with TS compilation:

- selection of observation time points (minutes to years)
- calendar problems
- length of the series
- changes in measurements methodology
- etc.

Some of these problems lead to a violation of **homogeneity**

Homogeneous TS

- the series reflects only the natural fluctuations of the studied element.
- External influences (such as a change in measurement methodology) are suppressed



Important role of EDA (**Explanatory Data Analysis**)

Explanatory Data Analysis (EDA)

EDA is used to investigate data sets and summarize their main characteristics. It helps determine how best to manipulate data sources to get the answers you need, making it easier to discover patterns, spot anomalies, test a hypothesis, or check assumptions. EDA often employ visualization methods.

EVA often leads to a **transformation**:

1) to meet the prerequisites for subsequent analysis
2) to highlight the signal we are searching (e.g. trend)

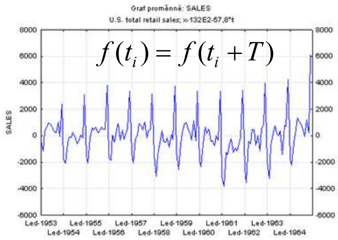
Common methods of transformation:

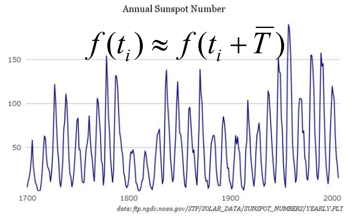
- add a constant — $y = y + C$
- linearization — $y = \ln(y)$
- remove a mean or trend — $y = y - \bar{y}$
- standardization (z-scores) — $y = \left(\frac{y - \bar{y}}{s_d} \right)$

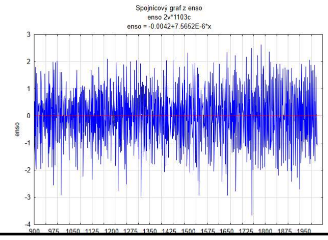
Time series components

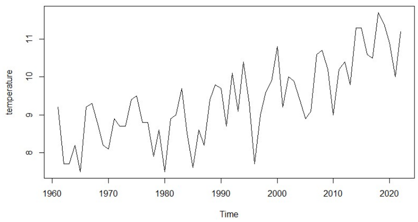
$Y_t = T_t + S_t + C_t + \varepsilon_t$

a) trend (T_t)
b) seasonality (S_t)
c) cyclicity (C_t)
d) random noise (ε_t)





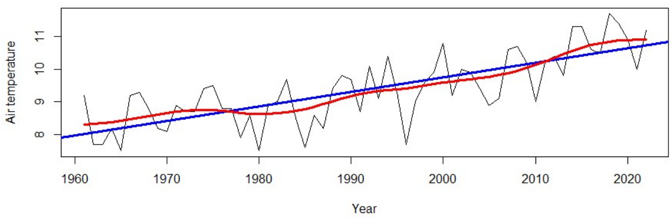




Time series trend

A trend is a general tendency of the development of the investigated phenomenon over a long period.

It is the result of long-term and permanent processes (on the scale of the assessed length of the time series).

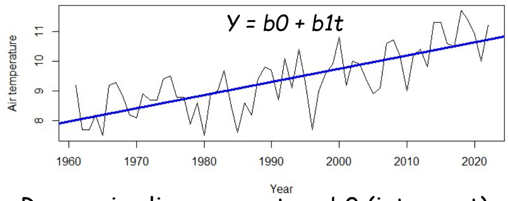


1) Monotonic trend – parameters are stable (e.g. linear regression) / deterministic methods
2) Adaptive trend – parameters change through time (e.g. moving averages)

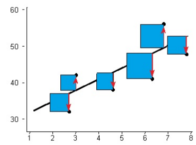
Trend analysis involves:

1) Choosing the appropriate type of trend
2) Trend parameters estimate
3) Testing statistical significance of trend parameters

Linear regression trend



Least squares method principle with residuals



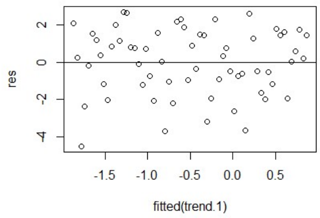
Regression line parameters b0 (intercept) and b1 (slope) are estimated by the **least squares method**

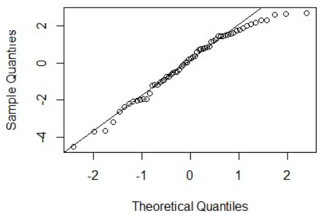
```
> trend.1 <- lm(bt$ANN ~ bt$YEAR)
> summary(trend.1)

Call:
lm(formula = bt$ANN ~ bt$YEAR)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.87025 -0.52417  0.05189  0.52439  1.17972 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -78.822528   9.972056  -7.904 7.14e-11 ***
bt$YEAR       0.044285   0.005007   8.844 1.81e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7055 on 60 degrees of freedom
Multiple R-squared:  0.5659,	Adjusted R-squared:  0.5587 
F-statistic: 78.22 on 1 and 60 DF,  p-value: 1.805e-12
```

Model verification

Analysis of residuals

```
> # plot resiuals
> res <- resid(trend.1)
> plot(fitted(trend.1), res)
> abline(0,0)
>
> #create Q-Q plot for residuals
> qqnorm(res)
> #add a straight diagonal line to the plot
> qqline(res)
```





Root Mean Square Error (RMSE)

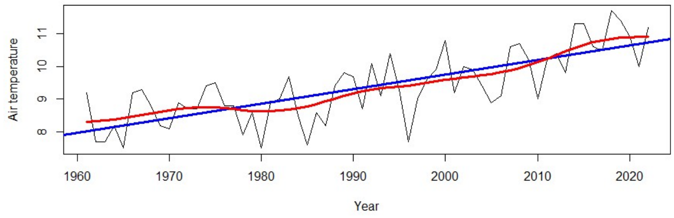
$$R.M.S.E. = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}}$$

```
> # calculate RMSE of regression model
> sqrt(mean(trend.1$residuals^2))
[1] 1.731824
> |
```

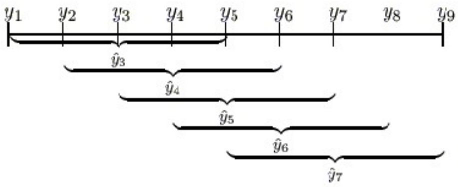
RMSE represents the average distance between the predicted values from the model and the actual values in the dataset.

Time series smoothing

It is used when the trend is changing and cannot be settled "globally" by a single mathematical curve (adaptive trend).



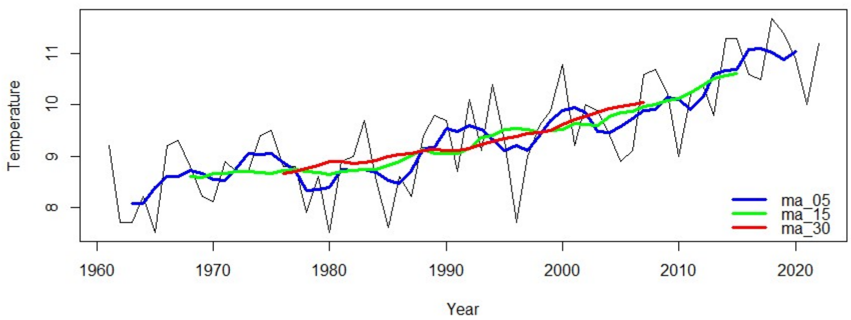
Moving averages



Smoothing methods highlight low-frequencies in time series (low-pass filters)

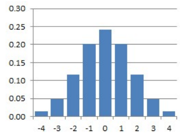
Moving averages

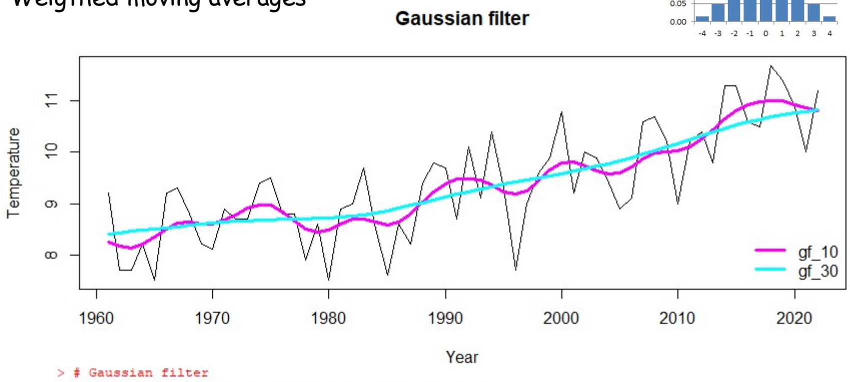
```
> x <- as.vector(bt$YEAR)
> y <- as.vector(bt$ANN)
> plot(x, y, type="l",xlab="Year", ylab="Temperature", main=" Moving averages")
>
> # Moving averages
> if(!require(forecast)){install.packages("forecast")}
> library(forecast)
>
> sm.l <- ma(y, order=5)
> lines(x,sm.l, col="blue",lwd=3)
> sm.l <- ma(y, order=15)
> lines(x,sm.l, col="green",lwd=3)
> sm.l <- ma(y, order=30)
> lines(x,sm.l, col="red",lwd=3)
>
> # Add legend
> legend(x="bottomright", legend=c("ma_05", "ma_15", "ma_30"), bty="n", lwd=3, col=c("blue", "green", "red"))
```



Gaussian filter

Weigthed moving averages






```
> # Gaussian filter
> if(!require(smoother)){install.packages("smoother")}
> library(smoother)
>
> plot(x, y, type="l",xlab="Year", ylab="Temperature", main="Gaussian filter")
>
> sm.2 <- smth.gaussian(y, window=10, tails=TRUE)
> lines(x,sm.2, col="magenta",lwd=3)
>
> sm.2 <- smth.gaussian(y, window=30, tails=TRUE)
> lines(x,sm.2, col="cyan",lwd=3)
>
> legend(x="bottomright", legend=c("gf_10", "gf_30"), bty="n", lwd=3, col=c("magenta", "cyan"))
```