Applied Time Series Analysis
FS 2012 – Week 01

Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

http://stat.ethz.ch/~dettling

ETH Zürich, February 20, 2012

Applied Time Series Analysis
FS 2012 – Week 01

Your Lecturer

Name:	Marcel Dettling
Age:	37 Years
Civil Status:	Married, 2 children
Education:	Dr. Math. ETH
Position:	Lecturer @ ETH Zürich and @ ZHAW Researcher in Applied Statistics @ ZHAW
Time Series:	Research with industry: *airlines*, *cargo*, *marketing* Academic research: *high-frequency financial data*

A First Example

In 2006, Singapore Airlines decided to place an order for new aircraft. It contained the following jets:

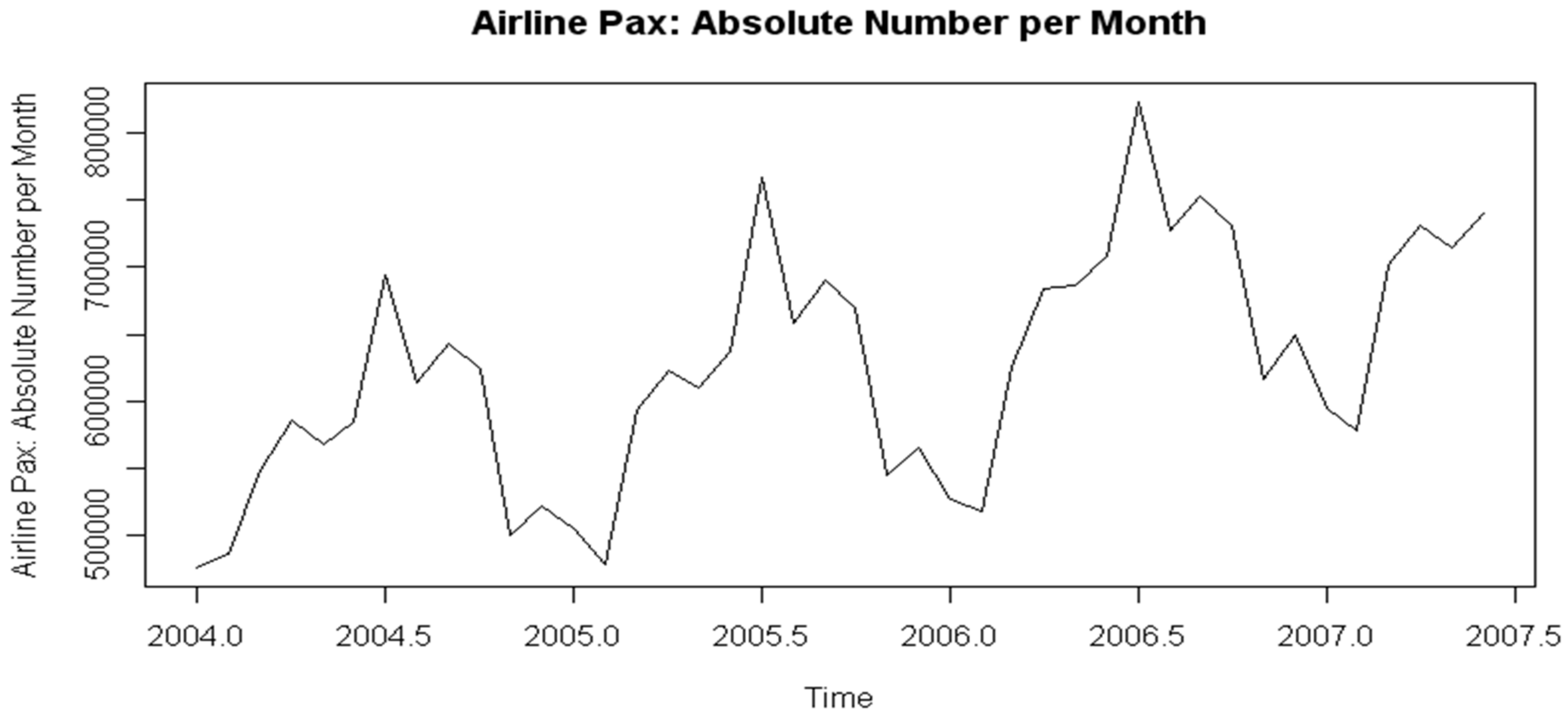
- 20 Boeing 787
- 20 Airbus A350
- 9 Airbus A380

How was this decision taken?

It was based on a combination of time series analysis on airline passenger trends, plus knowing the corporate plans for maintaining or increasing the market share.

A Second Example

- Taken from a former research project @ ZHAW
- Airline business: # of checked-in passengers per month



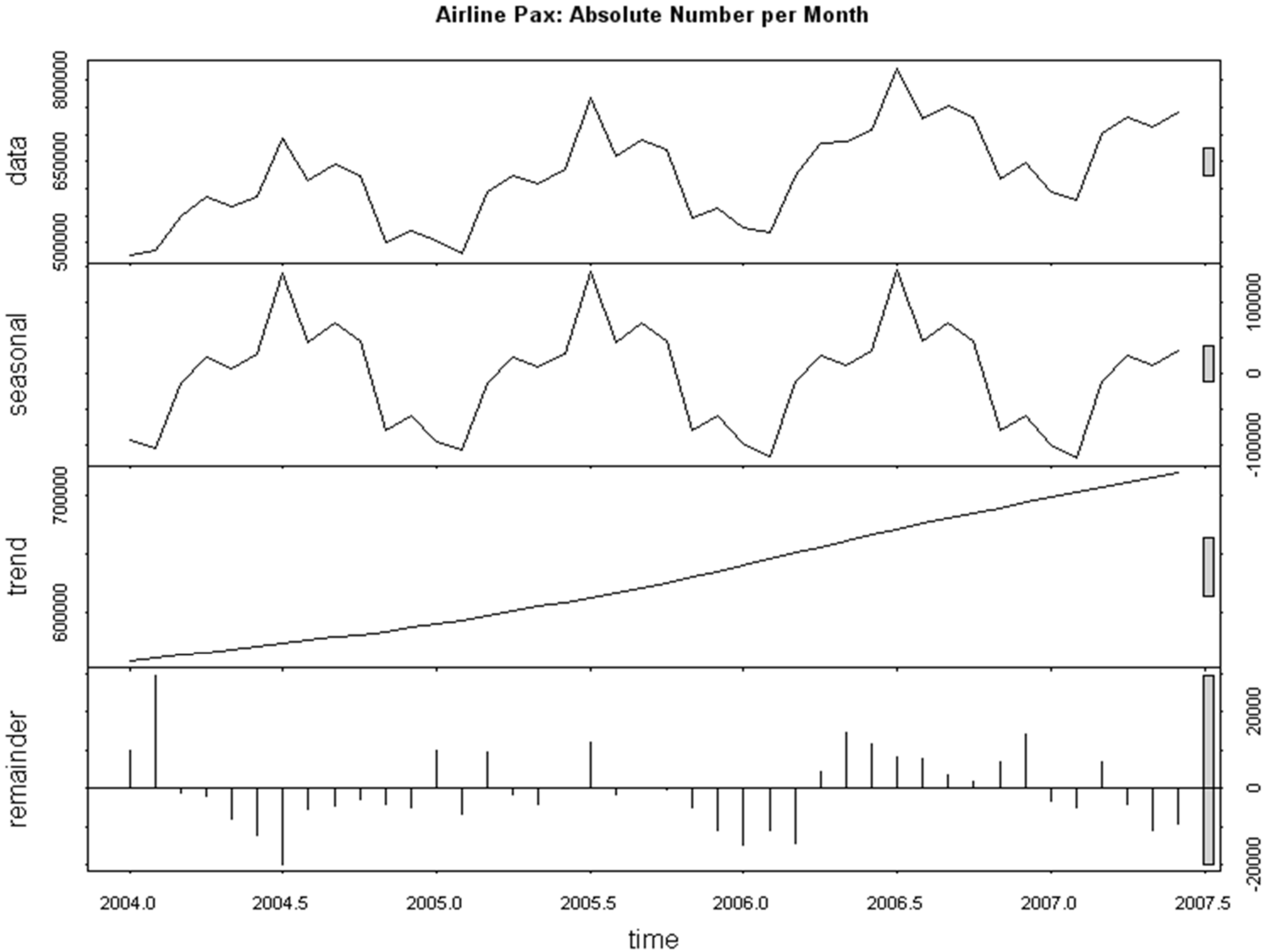
Some Properties of the Series

- Increasing trend (i.e. generally more passengers)
- Very prominent seasonal pattern (i.e. peaks/valleys)
- Hard to see details beyond the obvious

Goals of the Project

- Visualize, or better, extract trend and seasonal pattern
- Quantify the amount of random variation/uncertainty
- Provide the basis for a man-made forecast after mid-2007
- Forecast (extrapolation) from mid-2007 until end of 2008
- How can we better organize/collect data?

Applied Time Series Analysis
FS 2012 – Week 01



Organization of the Course

Contents:

- Basics, Mathematical Concepts, Time Series in R
- Descriptive Analysis (Plots, Decomposition, Correlation)
- Models for Stationary Series (AR(p), MA(q), ARMA(p,q))
- Non-Stationary Models (SARIMA, GARCH, Long-Memory)
- Forecasting (Regression, Exponential Smoothing, ARMA)
- Miscellaneous (Multivariate, Spectral Analysis, State Space)

Goal:

The students acquire experience in analyzing time series problems, are able to work with the software package R, and can perform time series analyses correctly on their own.

Applied Time Series Analysis
FS 2012 – Week 01

Organization of the Course

Applied Time Series Analysis – FS 2011

People:

Lecturer: **Dr. Marcel Dettling** (marcel.dettling@zhaw.ch)

Assistant: **Alain Hauser** (alhauser@ethz.ch)

Course Schedule:

All lectures and exercises will be held at HG E1.2, on Mondays from 10.15-11.55.

Week	Date	L/E	Topics
01	21.02.2011	L/L	Introduction to time series analysis
02	28.02.2011	L/E	Stationarity, decomposition of time series
03	07.03.2011	L/L	Autocorrelation, Correlogram
04	14.03.2011	L/E	Autoregressive Modeling 1
05	21.03.2011	L/L	Autoregressive Modeling 2
06	28.03.2011	L/E	Time series forecasting
07	04.04.2011	L/L	ARMA-Modeling 1
08	11.04.2011	L/E	ARMA-Modeling 2
09	18.04.2011	L/L	Time series regression
10	02.05.2011	L/E	Multivariate time series
11	09.05.2011	L/E	State space models
12	16.05.2011	L/L	Spectral Analysis 1
13	23.05.2011	L/E	Spectral Analysis 2
14	30.05.2011	L/L	Advanced Topics

Exercise Schedule:

The exercises will be held roughly every second week in the lecture room HG E1.2. There is only one group, for which an assistant will provide some background and useful hints on how to approach the problems.

Solving the problems needs to be done autonomously and requires the use of the statistical software package R. The exercise schedule is as follows:

Series	Date	Topic	Hand-In	Solutions
01	28.02.2011	Time series in R	07.03.2011	14.03.2011
02	14.03.2011	Autocorrelation	21.03.2011	28.03.2011
03	28.03.2011	AR(p)-modeling	04.04.2011	11.04.2011
04	11.04.2011	ARMA(p,q)-modeling	18.04.2011	02.05.2011
05	02.05.2011	Multivariate time series	09.05.2011	16.05.2011
06	09.05.2011	State space modeling	16.05.2011	23.05.2011
07	23.05.2011	Spectral analysis	26.05.2011	30.05.2011

The solved exercises can be handed in in the lectures where an assistant will pick them up. Sending them via e-mail or placing them in the corresponding tray in HG J68 until 11.55am of the due date is another option. Please write down your findings and comments. You can support this with a few plots, but please avoid handing in any R-code or an excessive amount of plots.

→ more details are given on the additional organization sheet

Introduction: What is a Time Series?

A time series is a set of observations x_t , where each of the observations was made at a specific time t.

- the set of times T is discrete and finite
- observations were made at fixed time intervals
- continuous and irregularly spaced time series are not covered

Rationale behind time series analysis:

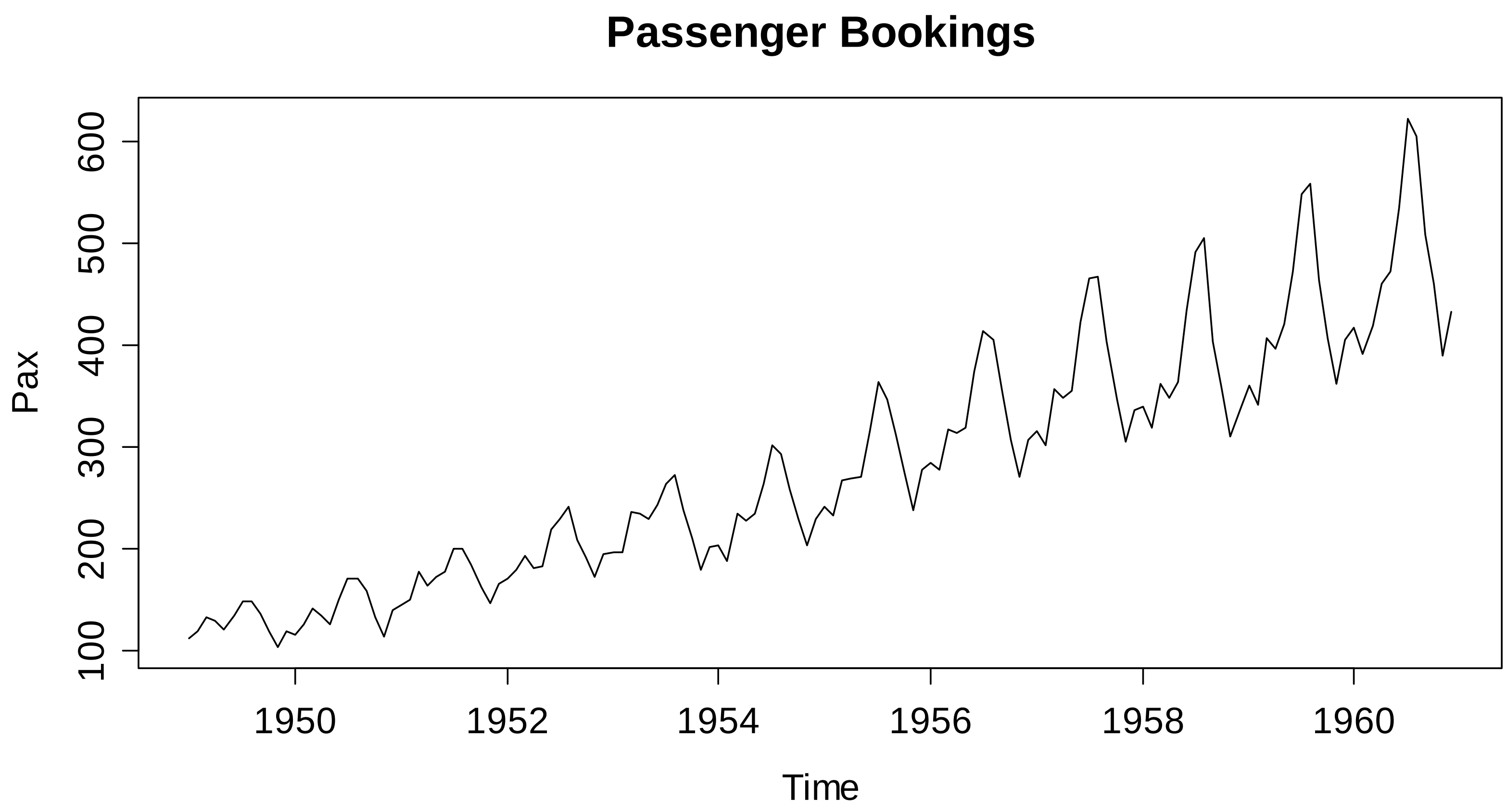
The rationale in time series analysis is to understand the past of a series, and to be able to predict the future well.

Example 1: Air Passenger Bookings

```
> data(AirPassengers)
> AirPassengers
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
```

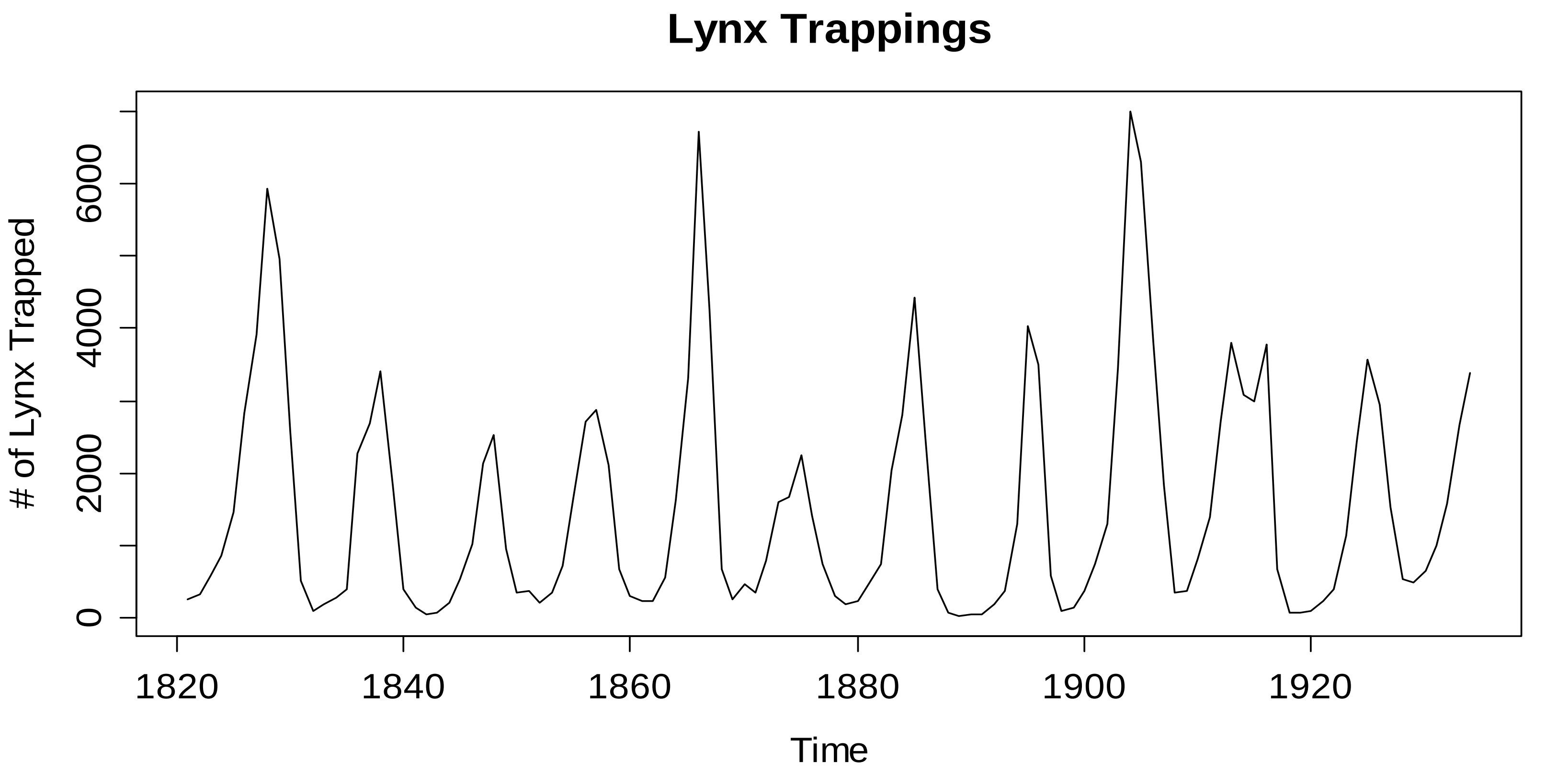
Example 1: Air Passenger Bookings

```
> plot(AirPassengers, ylab="Pax", main="Pax Bookings")
```



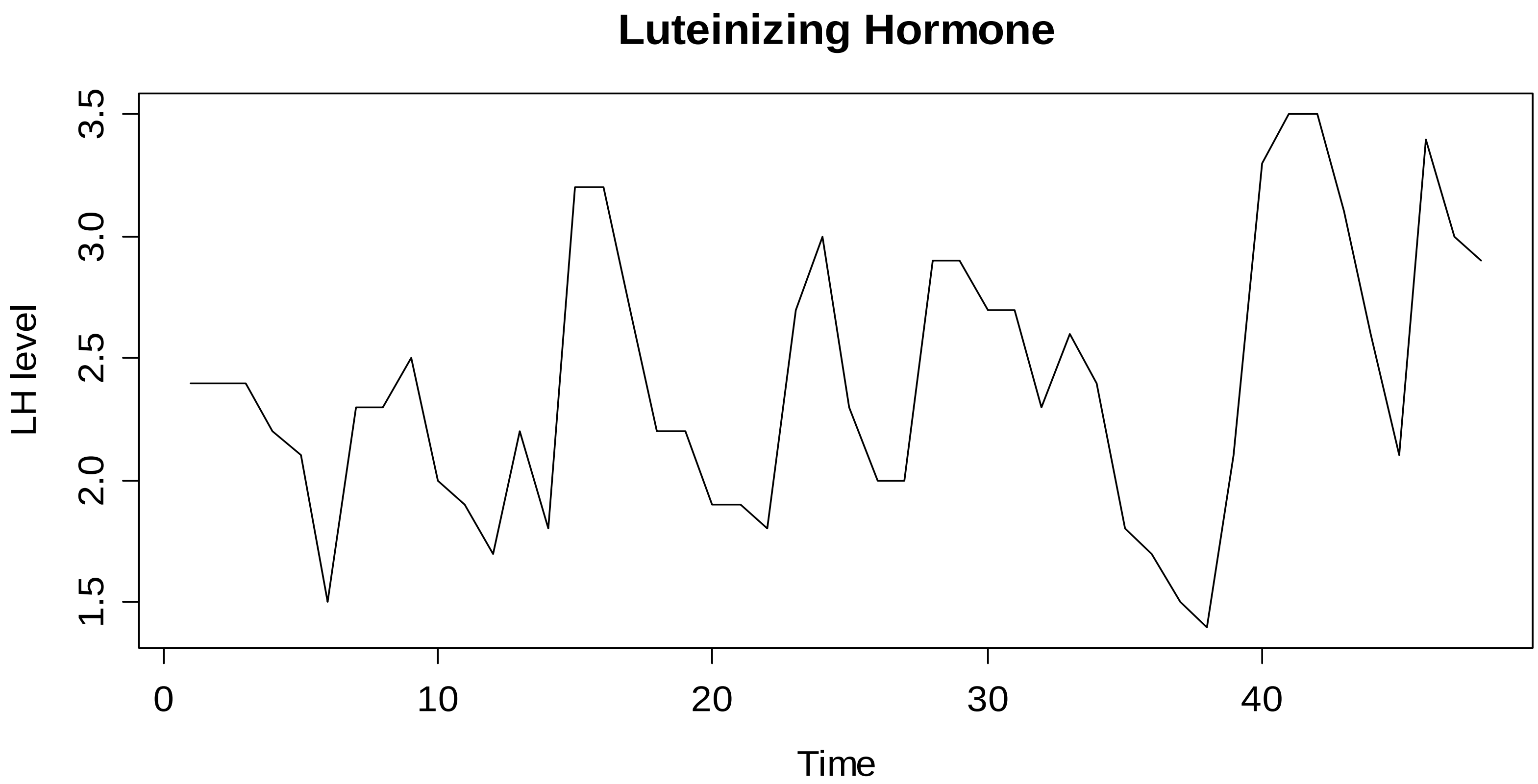
Example 2: Lynx Trappings

```
> data(lynx)
> plot(lynx, ylab="# of Lynx", main="Lynx Trappings")
```



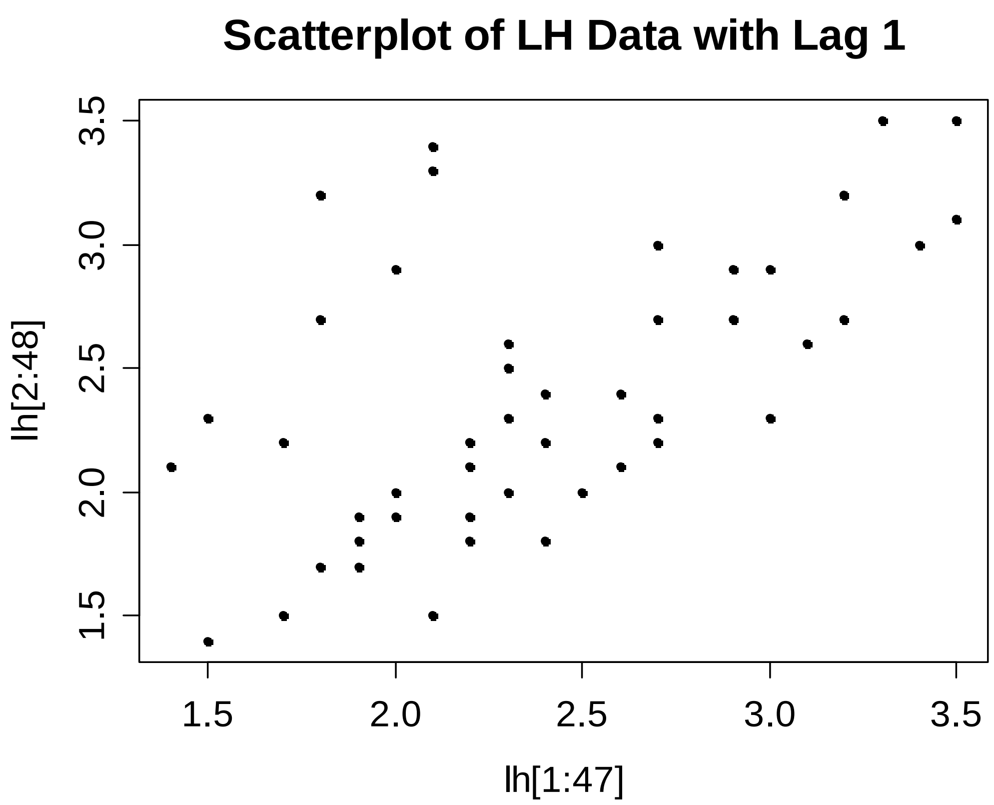
Example 3: Luteinizing Hormone

```
> data(lh)
> plot(lh, ylab="LH level", main="Luteinizing Hormone")
```



Example 3: Lagged Scatterplot

```
> plot(lh[1:47], lh[2:48], pch=20)
> title("Scatterplot of LH Data with Lag 1")
```



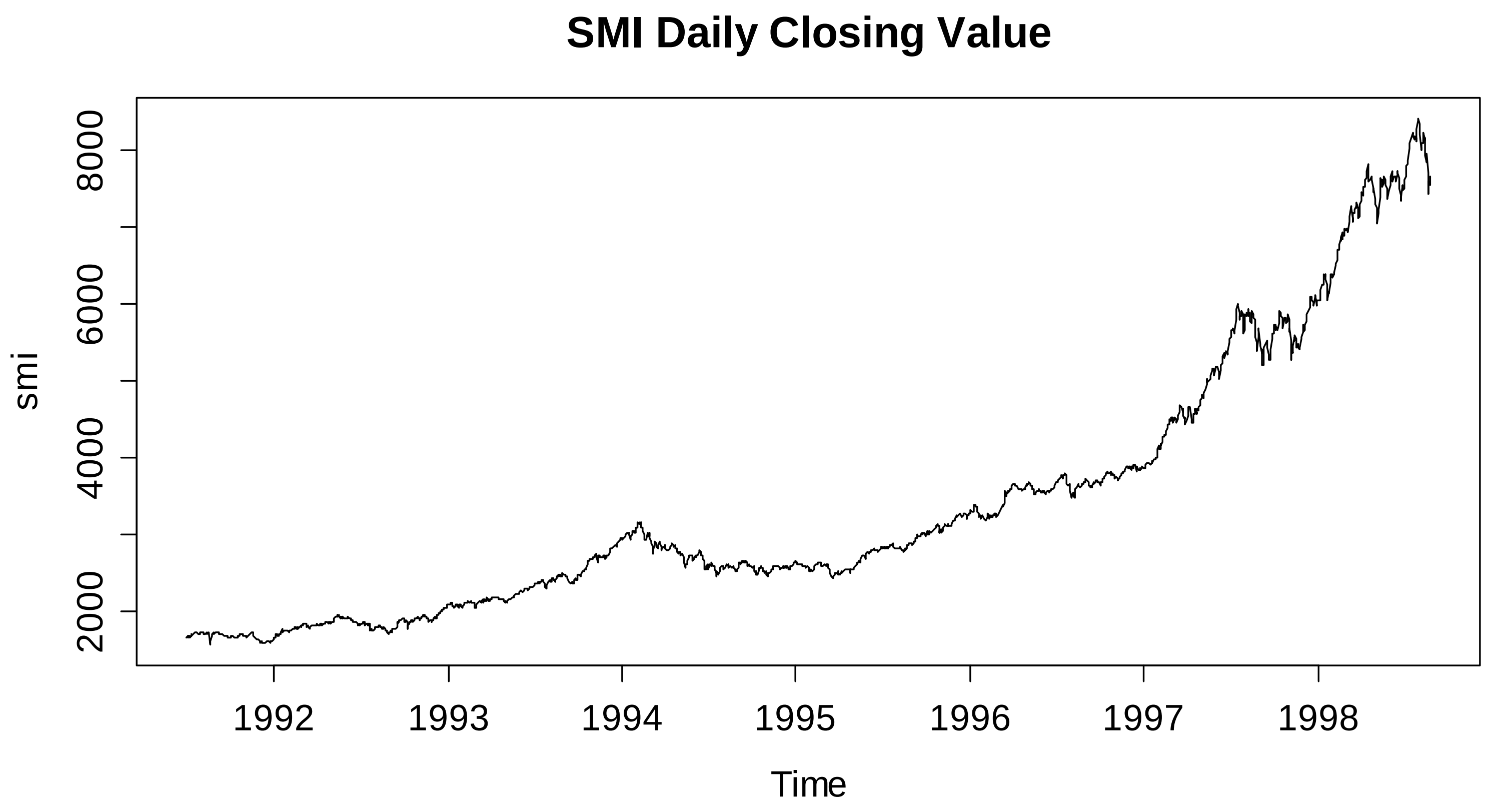
Example 4: Swiss Market Index

We have a multiple time series object:

```
> data(EuStockMarkets)
> EuStockMarkets
Time Series:
Start = c(1991, 130)
End = c(1998, 169)
Frequency = 260
             DAX    SMI    CAC   FTSE
1991.496 1628.75 1678.1 1772.8 2443.6
1991.500 1613.63 1688.5 1750.5 2460.2
1991.504 1606.51 1678.6 1718.0 2448.2
1991.508 1621.04 1684.1 1708.1 2470.4
1991.512 1618.16 1686.6 1723.1 2484.7
1991.515 1610.61 1671.6 1714.3 2466.8
```

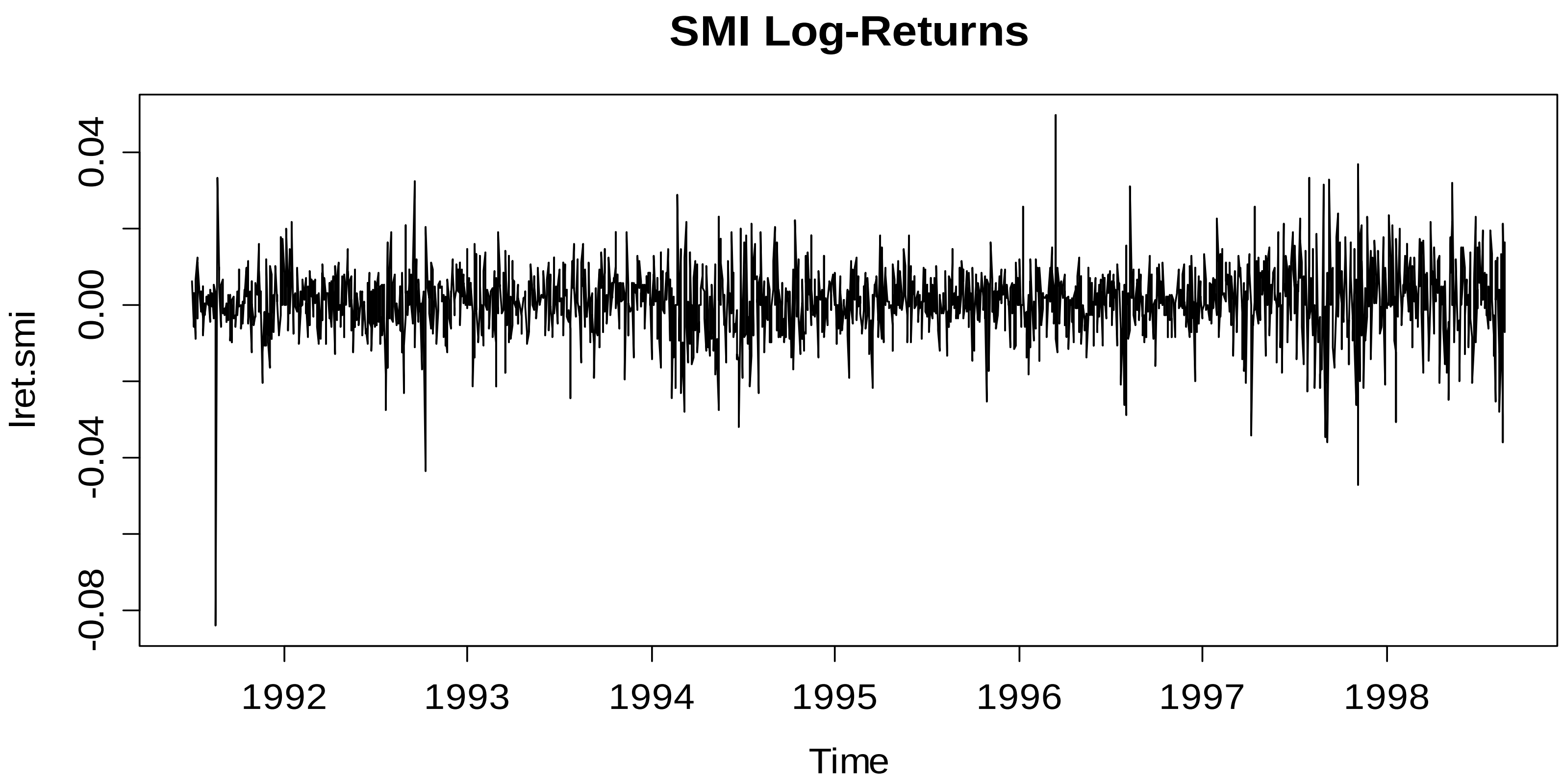
Example 4: Swiss Market Index

```
> smi <- ts(tmp, start=start(esm), freq=frequency(esm))
> plot(smi, main="SMI Daily Closing Value")
```



Example 4: Swiss Market Index

```
> lret.smi <- log(smi[2:1860]/smi[1:1859])
> plot(lret.smi, main="SMI Log-Returns")
```



Goals in Time Series Analysis

1) Exploratory Analysis

Visualization of the properties of the series

- *time series plot*
- *decomposition into trend/seasonal pattern/random error*
- *correlogram for understanding the dependency structure*

2) Modeling

Fitting a stochastic model to the data that represents and reflects the most important properties of the series

- *done exploratory or with previous knowledge*
- *model choice and parameter estimation is crucial*
- *inference: how well does the model fit the data?*

Goals in Time Series Analysis

3) Forecasting

Prediction of future observations with measure of uncertainty

- *mostly model based, uses dependency and past data*
- *is an extrapolation, thus often to take with a grain of salt*
- *similar to driving a car by looking in the rear window mirror*

4) Process Control

The output of a (physical) process defines a time series

- *a stochastic model is fitted to observed data*
- *this allows understanding both signal and noise*
- *it is feasible to monitor normal/abnormal fluctuations*

Goals in Time Series Analysis

5) Time Series Regression

Modeling response time series using 1 or more input series

$$Y_t = \beta_0 + \beta_1 u_t + \beta_2 v_t + E_t$$

where E_t is independent of u_t and v_t, but not i.i.d.

Example: $(\text{Ozone})_t = (\text{Wind})_t + (\text{Temperature})_t + E_t$

Fitting this model under i.i.d error assumption:

- leads to unbiased estimates, but...
- often grossly wrong standard errors
- thus, confidence intervals and tests are misleading

Stochastic Model for Time Series

Def: A *time series process* is a set $\{X_t, t \in \mathrm{T}\}$ of random variables, where T is the set of times. Each of the random variables $X_t, t \in \mathrm{T}$ has a univariate probability distribution F_t.

- If we exclusively consider time series processes with equidistant time intervals, we can enumerate $\{T = 1, 2, 3, ...\}$
- An observed time series is a realization of $X = (X_1, \ldots, X_n)$, and is denoted with small letters as $x = (x_1, \ldots, x_n)$.
- We have a multivariate distribution, but only 1 observation (i.e. 1 realization from this distribution) is available. In order to perform "statistics", we require some additional structure.

Stationarity

For being able to do statistics with time series, we require that the series “doesn’t change its probabilistic character” over time. This is mathematically formulated by **strict stationarity**.

Def: A time series $\left\{X_t, t \in \mathrm{T}\right\}$ is strictly stationary, if the joint distribution of the random vector $(X_t, \ldots, X_{t+k})$ is equal to the one of $(X_s, \ldots, X_{s+k})$ for all combinations of t, s and k.

➔

$X_t \sim F$	all X_t are identically distributed
$E[X_t] = \mu$	all X_t have identical expected value
$Var(X_t) = \sigma^2$	all X_t have identical variance
$Cov(X_t, X_{t+h}) = \gamma_h$	the autocov depends only on the lag h

Stationarity

It is impossible to „prove" the theoretical concept of stationarity from data. We can only search for evidence in favor or against it.

However, with strict stationarity, even finding evidence only is too difficult. We thus resort to the concept of *weak stationarity*.

Def: A time series $\{X_t, t \in \mathrm{T}\}$ is said to be *weakly stationary*, if

$$E[X_t] = \mu$$

$$Cov(X_t, X_{t+h}) = \gamma_h \quad \text{for all lags } h$$

and thus also: $Var(X_t) = \sigma^2$

Note that weak stationarity is sufficient for „practical purposes".

Testing Stationarity

- In time series analysis, we need to verify whether the series has arisen from a stationary process or not. Be careful: stationarity is a property of the process, and not of the data.

- Treat stationarity as a hypothesis! We may be able to reject it when the data strongly speak against it. However, we can never prove stationarity with data. At best, it is plausible.

- Formal tests for stationarity do exist (→ see scriptum). We discourage their use due to their low power for detecting general non-stationarity, as well as their complexity.

→ **Use the time series plot for deciding on stationarity!**

Evidence for Non-Stationarity

- **Trend**, i.e. non-constant expected value
- **Seasonality**, i.e. deterministic, periodical oscillations
- **Non-constant variance**, i.e. multiplicative error
- **Non-constant dependency structure**

Remark:

Note that some periodical oscillations, as for example in the lynx data, can be stochastic and thus, the underlying process is assumed to be stationary. However, the boundary between the two is fuzzy.

Strategies for Detecting Non-Stationarity

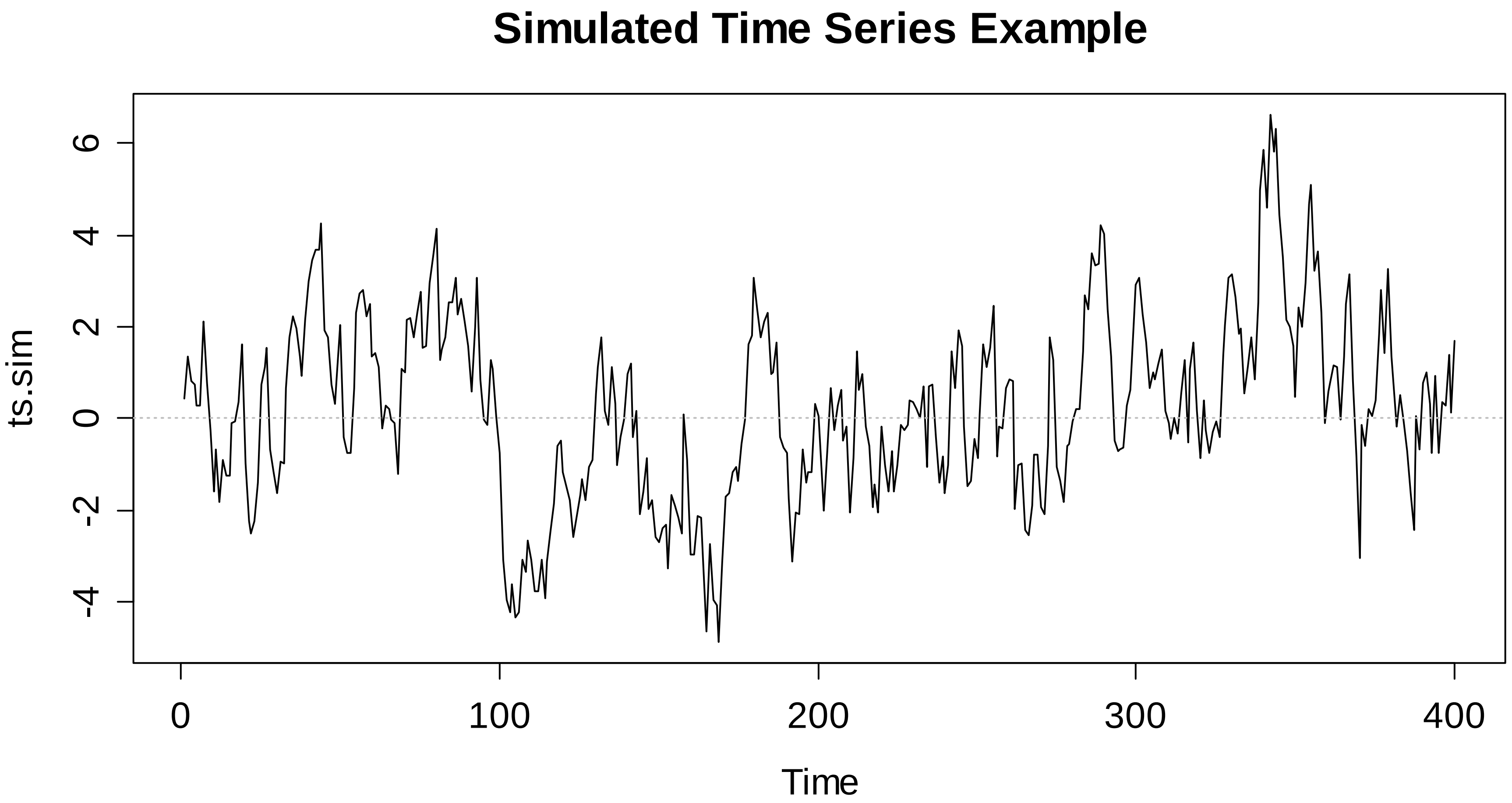
1) **Time series plot**
   - non-constant expected value (trend/seasonal effect)
   - changes in the dependency structure
   - non-constant variance

2) **Correlogram (presented later...)**
   - non-constant expected value (trend/seasonal effect)
   - changes in the dependency structure

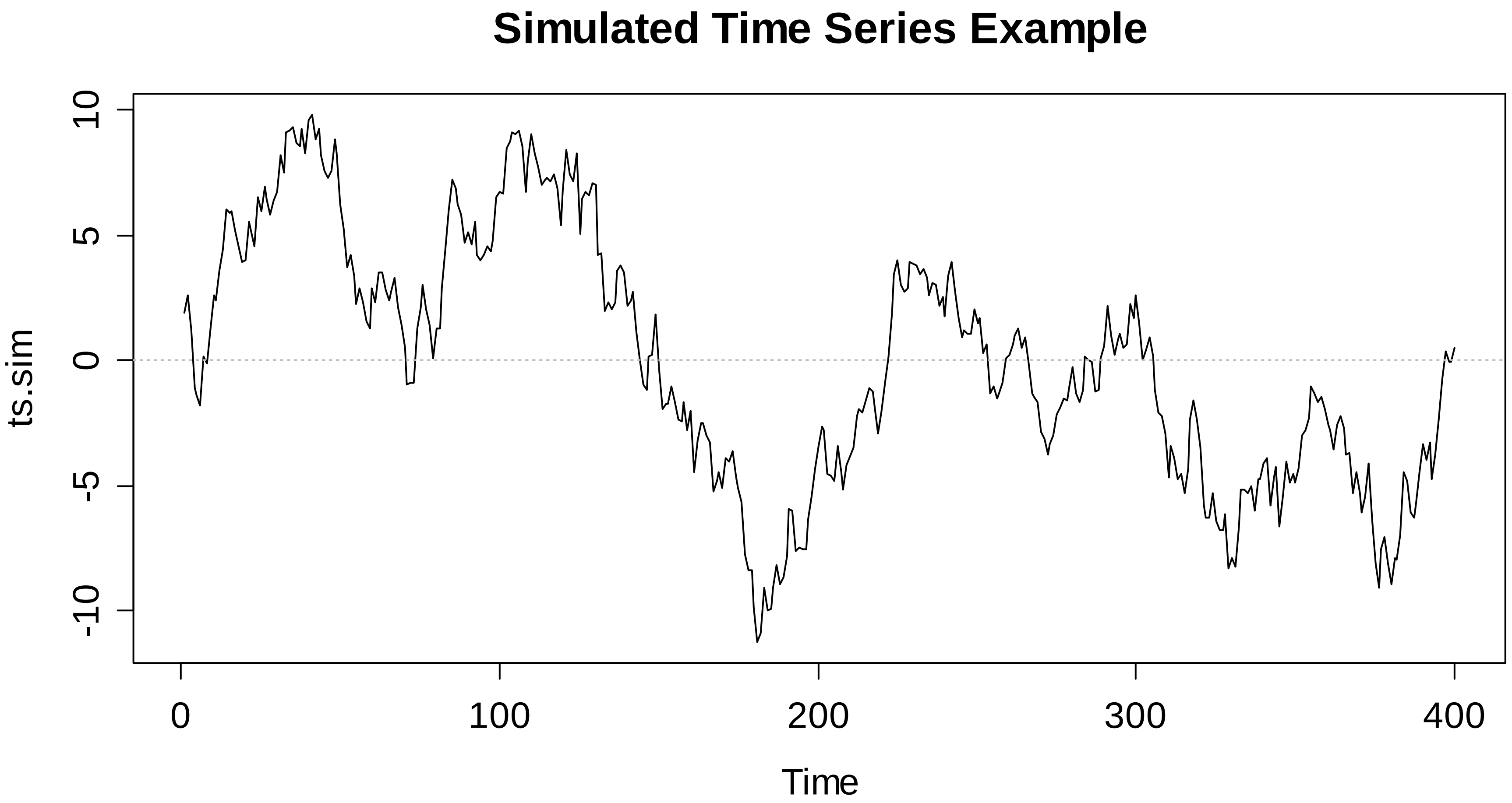
A (sometimes) useful trick, especially when working with the correlogram, is to split up the series in two or more parts, and producing plots for each of the pieces separately.

Applied Time Series Analysis
FS 2012 – Week 01

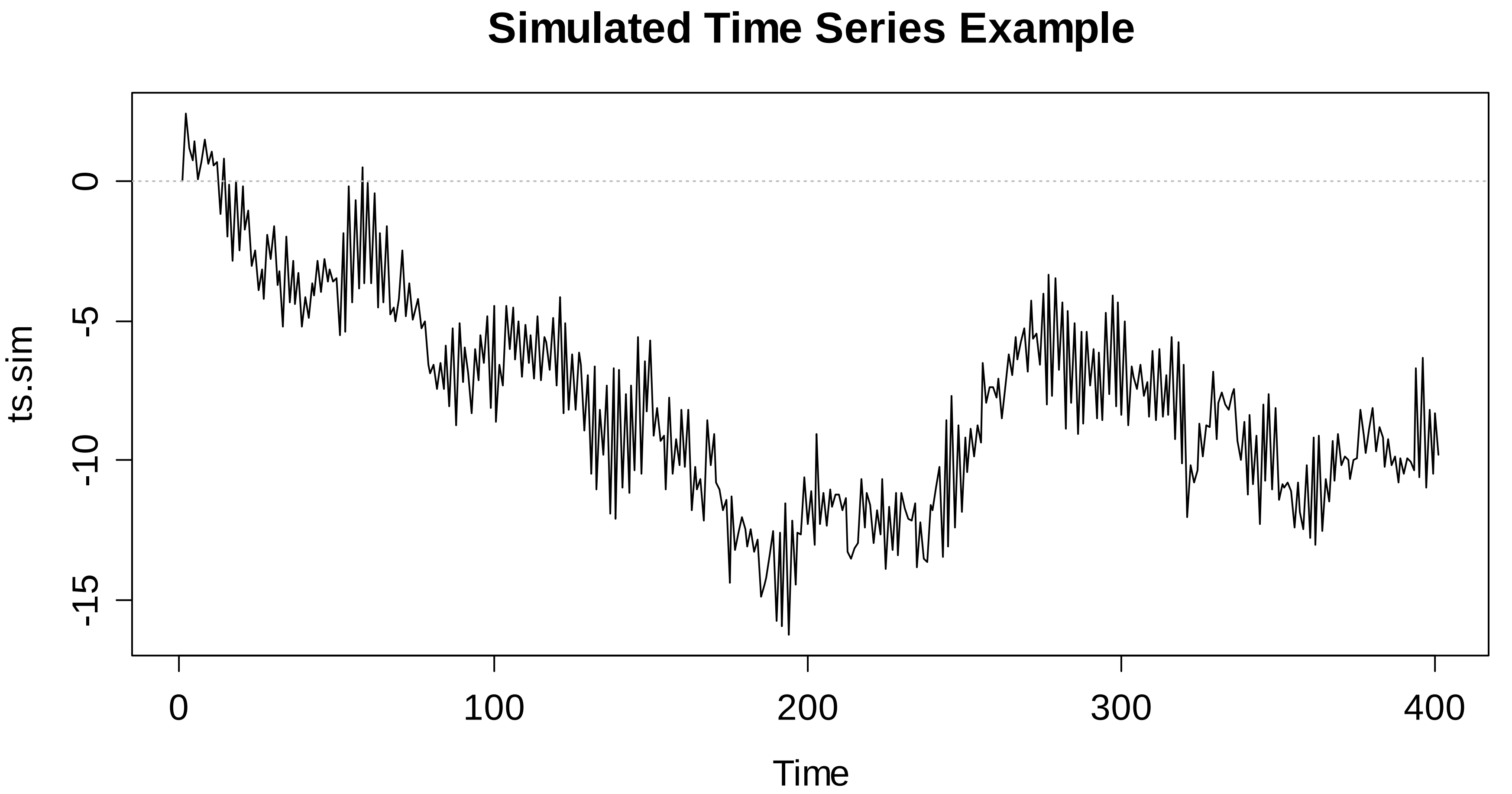
Example: Simulated Time Series 1



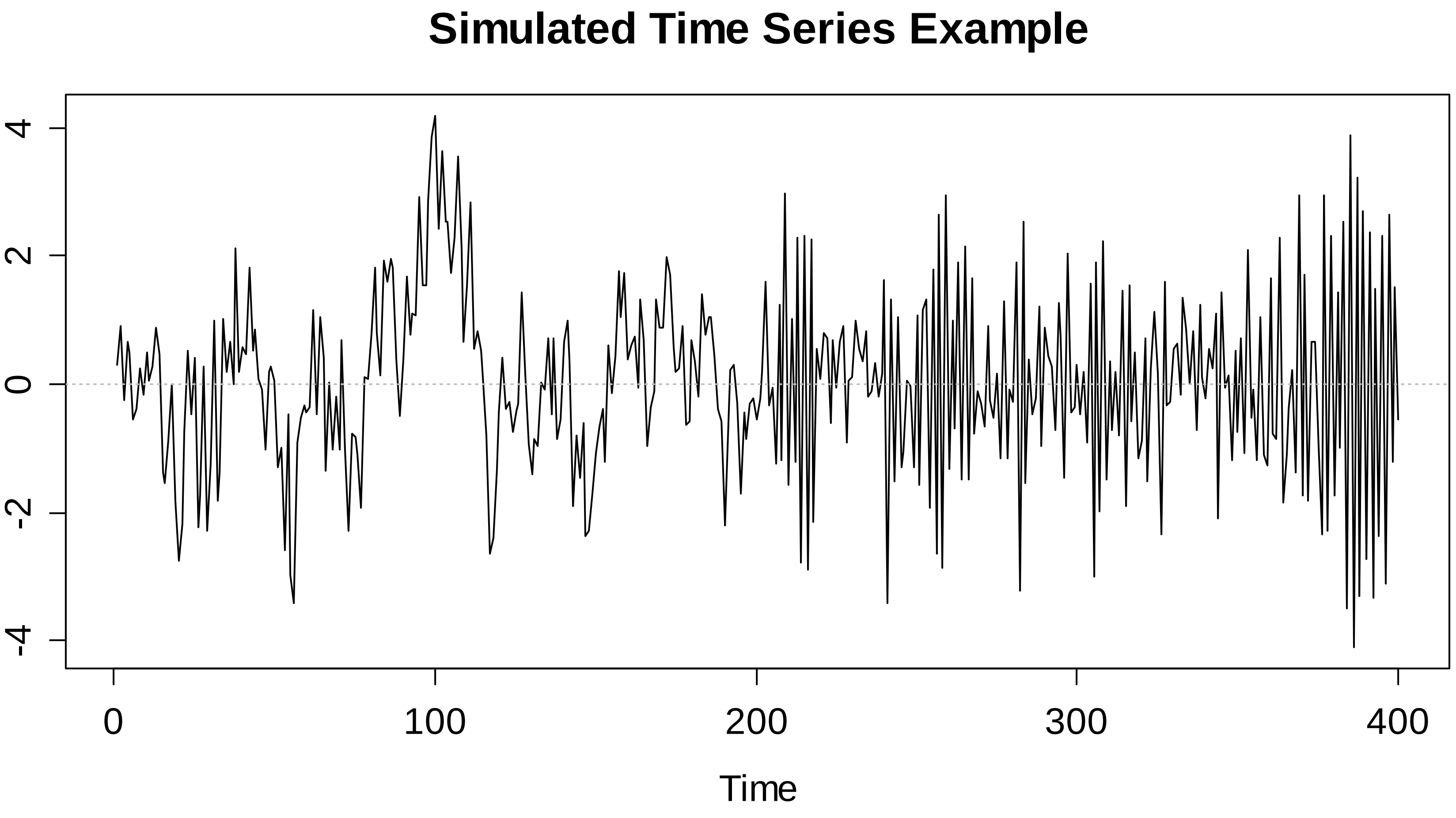
Example: Simulated Time Series 2



Example: Simulated Time Series 3



Example: Simulated Time Series 4



Time Series in R

- In **R**, there are *objects*, which are organized in a large number of *classes*. These classes e.g. include *vectors*, *data frames*, *model output*, *functions*, and many more. Not surprisingly, there are also *several classes for time series*.

- We focus on **ts**, the basic class for regularly spaced time series in **R**. This class is comparably simple, as it can only represent time series with *fixed interval records*, and *only uses numeric time stamps*, i.e. enumerates the index set.

- For defining a **ts** object, we have to supply the *data*, but also the *starting time* (as argument start), and the *frequency* of measurements as argument frequency.

Time Series in R: Example

Data: number of days per year with traffic holdups in front of the Gotthard road tunnel north entrance in Switzerland.

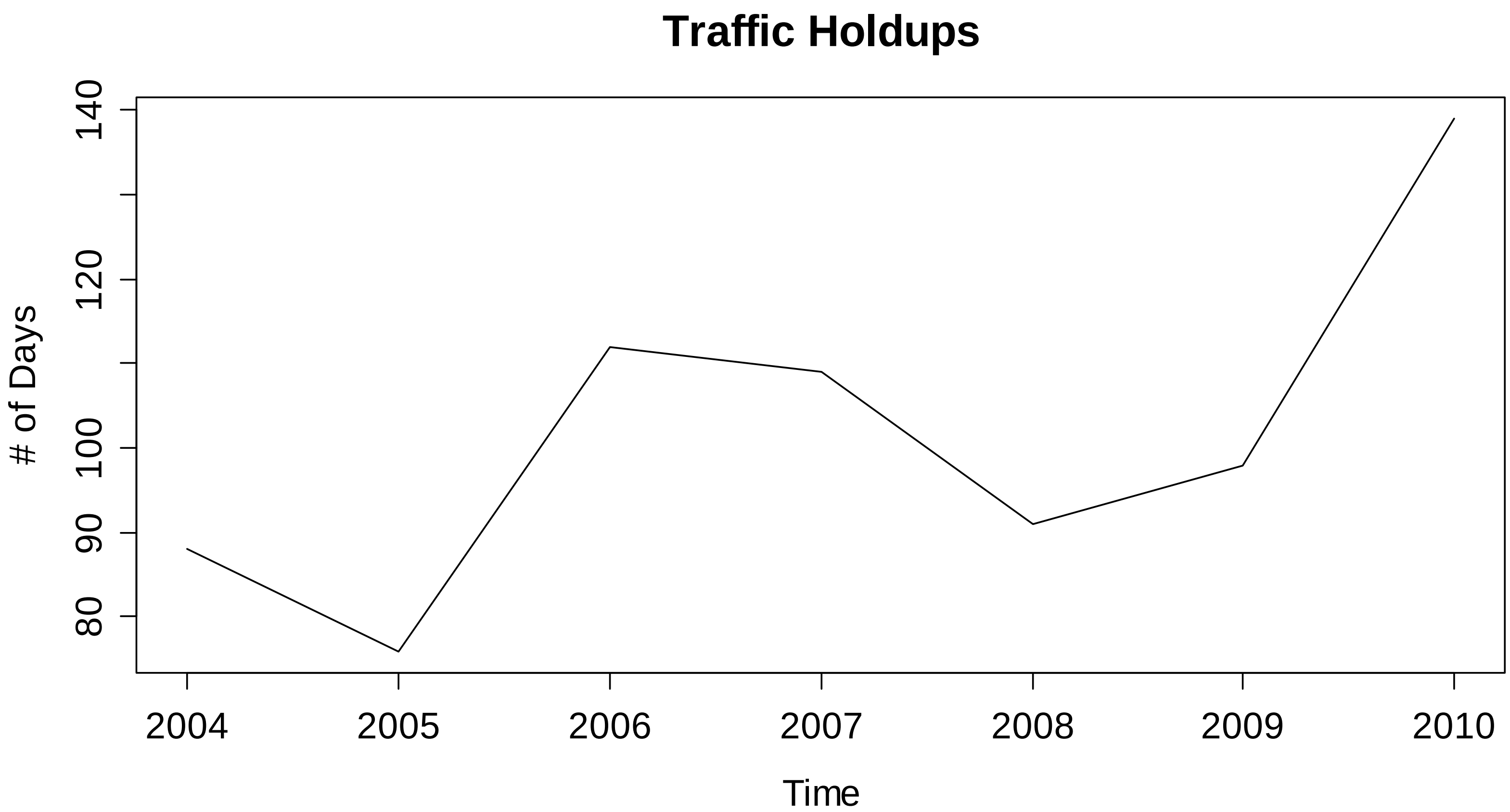
2004	2005	2006	2007	2008	2009	2010
88	76	112	109	91	98	139

```
> rawdat <- c(88, 76, 112, 109, 91, 98, 139)
> ts.dat <- ts(rawdat, start=2004, freq=1)

> ts.dat
Time Series: Start = 2004
End = 2010; Frequency = 1
[1]  88  76 112 109  91  98 139
```

Time Series in R: Example

```
> plot(ts.dat, ylab="# of Days", main="Traffic Holdups")
```



Further Topics in R

The scriptum discusses some further topics which are of interest when doing time series analysis in R:

- *Handling of dates and times in R*
- *Reading/Importing data into R*

→ **Please thoroughly read and study these chapters. Examples will be shown/discussed in the exercises.**