

# R - Einführung durch Angewandte Statistik (LVA 298.923)

## Block 2, Part 1: ggplot2 and doBy

**Ass.-Prof. Dr. Wolfgang Trutschnig**

FB Mathematik  
Universität Salzburg  
[www.trutschnig.net](http://www.trutschnig.net)

Salzburg, 2015-11-16/19



## What is this about ?

- ▶ Calculating simple aggregations as well as producing meaningful and pretty (!) graphics are usually the first steps in any data analysis
- ▶ We will have a first look at two packages that are particularly useful for that:
- ▶ **ggplot2**: Extremely flexible (and widely-used) package for the production of elegant graphics; Autor: Hadley Wickham (Entwicklung seit 2005), 'ggplot2: Elegant Graphics for Data Analysis' (Springer)
- ▶ **doBy**: Package to calculate 'groupwise summary statistics' (without loops) 'do something on data which is grouped **By** some variables'. Author: Soren Hojsgaard



## Learning by doing

- ▶ To illustrate functionality of the packages we will study one (almost real) dataset:
- ▶ ATM.txt, available at [www.trutschnig.net/courses](http://www.trutschnig.net/courses)

ymd	weekday	nr_weekday	sum_out	holiday
2007-01-01	Mon	1	4040	1.00
2007-01-02	Tue	2	22760	1.50
2007-01-03	Wed	3	18810	0.00
2007-01-04	Thu	4	24910	0.00
2007-01-05	Fri	5	25650	0.50
2007-01-06	Sat	6	5650	1.00

- ▶ The data contains (almost real) timeseries (3 years) of daily withdrawn amounts at an ATM in northern Spain.
- ▶ Original project statement: Develop reliable forecasts for daily withdrawn amounts in order to optimize the refill (cash-delivery) system (500 ATMs).

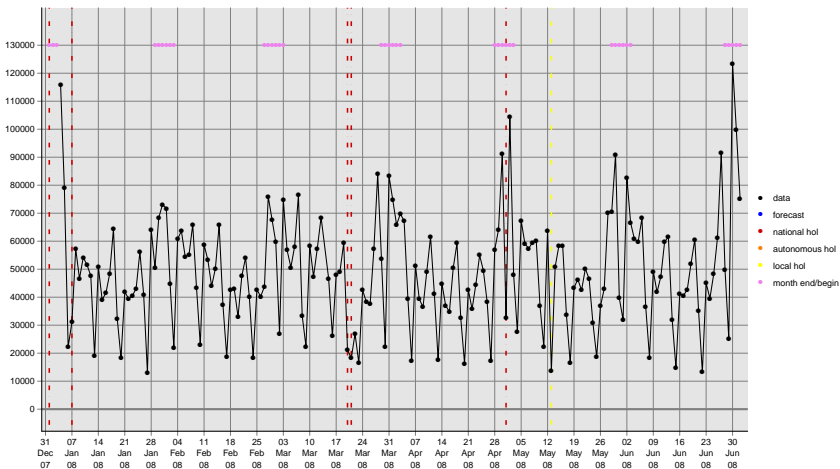


We analyze the data using ggplot2 and doBy proceeding in the following steps (Code snippets can be found at [www.trutschnig.net/courses](http://www.trutschnig.net/courses)):

1. Data import and a first glance at the data.
2. Pretty plot of the timeseries - can any patters be found ?
3. Yearly histograms and boxplots - patterns confirmed ?
4. More sophisticated boxplots (violins) per weekday and day of the month - patterns ?
5. Influence of public holidays ?
6. Export/saving of nice plots.
7. How to forecast the withdrawn amounts ?
8. Calculate summary statistics via doBy, compare with graphics



15day forecast unit 2  
 Center: 0; Number Machines: 4  
 mean daily withdrawn amount: 40878  
 mean daily abs error: 6179, mean daily error to mean withdrawn: 15.12%



15day forecast unit 2  
 Center: 0; Number Machines: 4  
 mean daily withdrawn amount: 40878  
 mean daily abs error: 6179, mean daily error to mean withdrawn: 15.12%

