

# Wissenschaftliches Rechnen 405.100

## knitR: learning by doing

**Univ.-Prof. Dr. Wolfgang Trutschnig**

Arbeitsgruppe Stochastik/Statistik  
Fachbereich Mathematik  
Universität Salzburg  
[www.trutschnig.net](http://www.trutschnig.net)

Salzburg, WS 2021/22



## Starting point

- ▶ Very common situation (in research and in business):
- ▶ An elaborate report including graphics, tables and text has been produced (using Excel, Word, Powerpoint etc.)
- ▶ New data arrives, often on a weekly/monthly basis, or new data is added to an existing dataset
- ▶ The same report is needed for the new/updated data
- ▶ **Do all calculations again, prepare all the tables and graphics again, and reproduce the report?**
- ▶ Apart from wasting time and money, which other problems occur?
  - ▶ Copy and paste is dangerous...
  - ▶ Small (copy and paste) mistakes can have far-reaching consequences
  - ▶ Can you assure to execute exactly the same manual data analysis steps as last time (Excel, etc.)?
- ▶ Reproducible results?



## A possible solution - knitr<sup>1</sup>

- ▶ Combine the power and flexibility of R with the typesetting capabilities of LaTeX
- ▶ R-Studio (user-interface for R) supports knitr
- ▶ **R, R-Studio and LaTeX are freeware** and run on all standard platforms!
- ▶ Let knitr do all the calculations, generate the graphics and tables and include numbers into the text
- ▶ Invest time in the first creation of the report - save time in all subsequent runs
- ▶ Use the saved time for more important and less annoying things than copy and paste (which robots can do much better than we)...

---

<sup>1</sup>or, alternatively, R Markdown



## Plan for today:

- ▶ Get knitR running on your laptops
- ▶ Run a minimal example
- ▶ Understand the basic building blocks of knitR
- ▶ Try to understand a pre-prepared knitR report
- ▶ Adjust/Manipulate/Extend the pre-prepared knitR report

**knitR has to be learned hands-on!**

- ▶ Download and install the necessary software: (Basic) Miktex, Texmaker, R, R-Studio (in this order)



- ▶ Each knitr-file has the extension **.Rnw**
- ▶ Each knitr-file consists of **two building blocks**:
  1. LaTeX code
  2. R-Code
- ▶ There are **two types of R-Code**:
  1. **chunks**, i.e. code as separate paragraph (produce tables, graphics, etc.)  
Each chunk starts with `<<some options>>=` and ends with `@`
  2. inline code, using the command `Sexpr`
- ▶ Knowing (basic) LaTeX and R is enough - the rest is combining blocks
- ▶ We start with a first minimal example demonstrating the LaTeX & R structure



```

1 \documentclass{article}
2 \begin{document}
3
4
5 <<histo,fig.width=10,fig.height=6,fig.cap='Histogram',echo=FALSE
  >>=
6 x<-rnorm(100,0,1)
7 hist(x,probability = TRUE,col="lightblue")
8 @
9
10 The estimates for  $\mu$  and  $\sigma$  are given by  $\overline{\{x\}}_n = \text{Sexpr}\{\text{mean}(x)\}$ 
11 and  $s^2_n = \text{Sexpr}\{\text{sd}(x)\}$ , whereby  $n = \text{Sexpr}\{\text{length}(x)\}$ .
12
13 \end{document}

```

knitr\_mini.Rnw

- ▶ Which part is LaTeX, which part is R?
- ▶ Download the file knitr\_mini.Rnw from [www.trutschnig.net/courses](http://www.trutschnig.net/courses), save it in a new folder, and open it with R-Studio



- ▶ A button **Compile PDF** appears
- ▶ Clicking it has the following effect
- ▶ knitr 'translates' the R-Codes into standard LaTeX code and produces a .tex-file
- ▶ Plots produced are, by default, saved as pdfs in a folder named 'figure'
- ▶ the .tex file is then compiled (pdfLaTeX) and a final pdf is produced
- ▶ All files are saved in the folder where the knitr-file is located



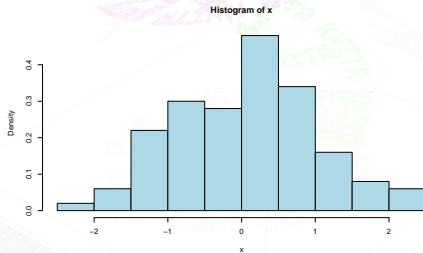


Figure 1: Histogram

The estimates for  $\mu$  and  $\sigma$  are given by  $\bar{x}_n = 0.0479857$  and  $s_n^2 = 0.9382978$ , whereby  $n = 100$ .





## Learning by doing - exercise(s)

### Exercise 1:

- ▶ Manipulate/Extend knitr\_mini.Rnw in such a way that the resulting pdf looks like knitr\_mini\_extended.pdf  
Hint: `par(mfrow = c(2,1))` can be used to have two plots in one graphic



```
1 %example loads the ATM data set, aggregates and produces a small  
2 summary  
3  
4 %Block 1: basic LaTeX settings  
5  
6 \documentclass[12pt]{article}  
7 \usepackage{amsmath}  
8 \usepackage{graphicx}  
9 \usepackage{hyperref}  
10 \usepackage{eurosym}  
11 \usepackage{color}  
12 \usepackage{float}  
13 \setlength{\textwidth}{650pt}  
14 \setlength{\textwidth}{480pt}  
15 \hoffset = -15mm  
16  
17 \begin{document}
```

knitr\_ATM/knitr\_ATM.Rnw



```
1
2 %Block 2: basic R setup: load all required packages
3 <<setup, include=FALSE, cache=FALSE>>=
4 library(knitr)
5 library(ggplot2)
6 library(doBy)
7 library(gridExtra)
8 library(RColorBrewer)
9 library(xtable)
10 Sys.setlocale("LC_TIME", "English")    #set English
11 # set global chunk options
12 opts_chunk$set(fig.path='figure/graphic-', fig.align='center', fig.
13     pos='!ht', echo=FALSE, warning = FALSE)
14     #global options for produced figures (name of figures,
15     automatic centering, etc.)
16     #echo=FALSE: don't include R-Code in output,
17     #warning=FALSE: print warning in console but not in pdf
18     #fig.pos='!ht': place it here in the doc
19 a<-Sys.time()
20 @
```

knitr\_ATM/knitr\_ATM.Rnw



```
1 \title{\vspace{-4cm} ATM withdrawals\footnote{This report was
   created on \Sexpr{a}}}  
2 \author{Wolfgang Trutschnig}  
3   %The footnote prints the acutal time calculated in the chunk  
4   above  
5 \maketitle  
6  
7 \section{Quick overview}
```

knitr\_ATM/knitr\_ATM.Rnw



```

1 <<results='asis'>>=
2 #Download ATM.txt and include table of first six rows in output
3 A<-read.table("http://www.trutchnig.net/ATM.txt",head=TRUE)
4 #A <- read.table("C:/Users/truts/Documents/Miktex_docs/Forschung und
      Uni/Skripten/Wissenschaftliches Rechnen/ATM.txt",head=TRUE)
5 A$ymd<-as.Date(A$ymd)
6 A$month<-as.numeric(substr(A$ymd,6,7))
7 A$year<-substr(A$ymd,1,4)
8 beg<-min(A$ymd); end<-max(A$ymd)
9 mis<-nrow(subset(A,is.na(A$sum_out)==1))
10 H<-subset(A,A$holiday==1)
11 V<-subset(A,A$holiday==0.5)
12 B<-A[1:6,1:5]
13 B$ymd<-as.character(B$ymd)
14 print(xtable(B,label="taba",caption="First six lines of the dataset
      "),size="footnotesize",include.rownames=FALSE)
15 #NB: xtable prepares table output for LaTeX

```

knitr\_ATM/knitr\_ATM.Rnw



- 1
- 2 The dataset (see Table `\ref{taba}`) contains daily withdrawn amounts in the period from `\Sexpr{beg}` till `\Sexpr{end}` (`\Sexpr{mis}` entries are missing).
- 3 We expect weekdays and holidays to have a strong influence on withdrawn amounts and, additionally, to see an impact of the financial crisis starting with autumn 2008.

knitR\_ATM/knitR\_ATM.Rnw



```
1
2 <<boxplot,fig.width=13, fig.height=6, fig.cap=paste('Boxplot per day
3   of week and year, the medians are also printed in Table \\ref{
4     tabb}')>>=
5   #size of the plot (width and height) in inches, NB: default output
6     width is \textwidth
7 p <- ggplot(data=A, aes(x=factor(nr_weekday), y=sum_out, fill=factor(nr
8   _weekday)))
9 p <- p + geom_boxplot(outlier.size=0)
10 p <- p + facet_wrap(~year)
11 p <- p + xlab("weekday")
12 p <- p + scale_fill_discrete(name = "Weekday")
13 p <- p + geom_point(data=H, colour="red", size=1.5)
14 p <- p + geom_point(data=V, colour="blue", size=1.5)
15 p <- p + theme_bw()
16 p
17 #NB: label of figure is automatically generated as fig:boxplot
```

knitr\_ATM/knitr\_ATM.Rnw



```
1
2 <<results='asis '>>=
3 medna<-function(x){median(x[is.na(x)==0])}
4   #little function calculating the median (ignoring missing values)
5 BB<-summaryBy(data=A,sum_out~nr_weekday,FUN=c(medna))
6   #calculate median withdrawn amount per weekday
7 WD<-A[1:7,2:3]
8 BB<-merge(BB,WD)
9 names(BB)[2]<-"sum_out"
10 BB$sum_out<-round(BB$sum_out)
```

knitr\_ATM/knitr\_ATM.Rnw





1

2 Considering all years together the median withdrawn amount is  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[1]\}$   $\backslash$ euro $\{\}$  on Mondays,  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[2]\}$   $\backslash$   
 $\text{euro}\{\}$  on Tuesdays,  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[3]\}$   $\backslash$   
 $\text{euro}\{\}$  on Wednesdays,  
 3  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[4]\}$   $\backslash$   
 $\text{euro}\{\}$  on Thursdays,  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[5]\}$   $\backslash$   
 $\text{euro}\{\}$  on Fridays,  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[6]\}$   $\backslash$   
 $\text{euro}\{\}$  on Saturdays,  
 and  $\backslash$   
 $\text{Sexpr}\{\text{BB}\$sum\_out[7]\}$   $\backslash$   
 $\text{euro}\{\}$  on Sundays.

knitR\_ATM/knitR\_ATM.Rnw



```

1
2 <<results='asis'>>=
3 AA<-summaryBy(data=A,sum_out~year+nr_weekday,FUN=c(medna)) #dplyr:
   AA <- A %>% group_by(year, nr_weekday) %>% dplyr::summarise(sum_
   out=medna(sum_out))
4 WD<-A[1:7,2:3]
5 AA<-merge(AA,WD)
6 AA<-AA[order(AA$nr_weekday,AA$year),]
7 names(AA)[3]<-"sum_out"
8 AA<-subset(AA,select=c(weekday,year,sum_out))
9 print(xtable(AA,label="tabb",caption="Median withdrawn amount per
   year and day of week"),size="scriptsize",include.rownames=FALSE)

```

knitr\_ATM/knitr\_ATM.Rnw



```
1
2 <<boxplot_monthly, fig.width=13, fig.height=6, fig.cap='Boxplot per
   month, the impact of the financial crisis starting with autumn
   2008 '>>=
3 B<-A
4 B$year<-as.factor(A$year)
5 B$month<-as.factor(A$month)
6 farben<-c("gray50", "magenta", "green")
7 p <- ggplot(data=B, aes(x=month, y=sum_out))
8 p <- p + geom_boxplot(aes(fill=year), outlier.size=0)
9 p <- p + scale_fill_manual(values=farben)
10 p <- p + geom_jitter(colour="gray30")
11 p <- p + theme_bw()
12 p
13 @
```

knitr\_ATM/knitr\_ATM.Rnw



## Exercise 2:

- ▶ Produce a nice knitR report for the RTR dataset. The report should include (minimum requirements):
- ▶ At least three graphics (use the ones you already have from previous exercises)
- ▶ At least three tables (including some summary statistics)
- ▶ Some text informing about what the RTR-dataset is about
- ▶ Some values calculated from the data included in the text via `Sexpr`.

