HOST
Hochschule Stralsund

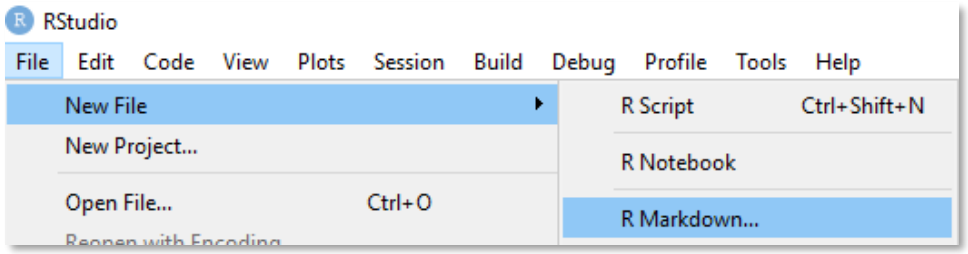
# Data Science

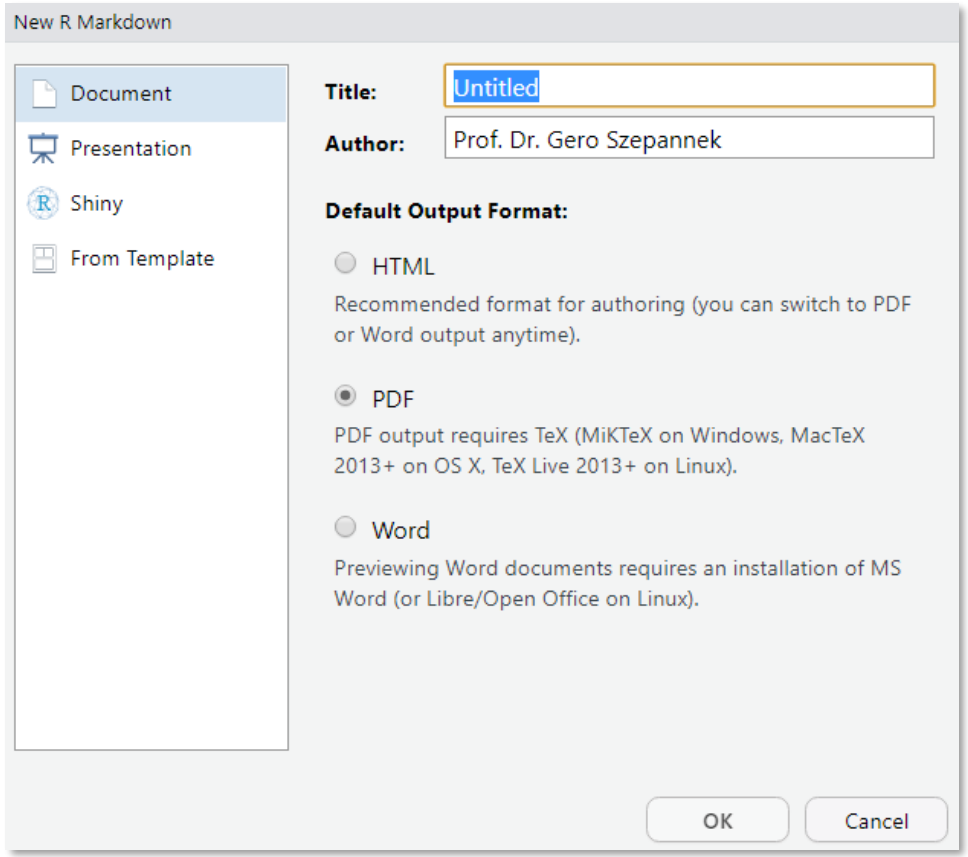
## Turning Analyses directly into Results

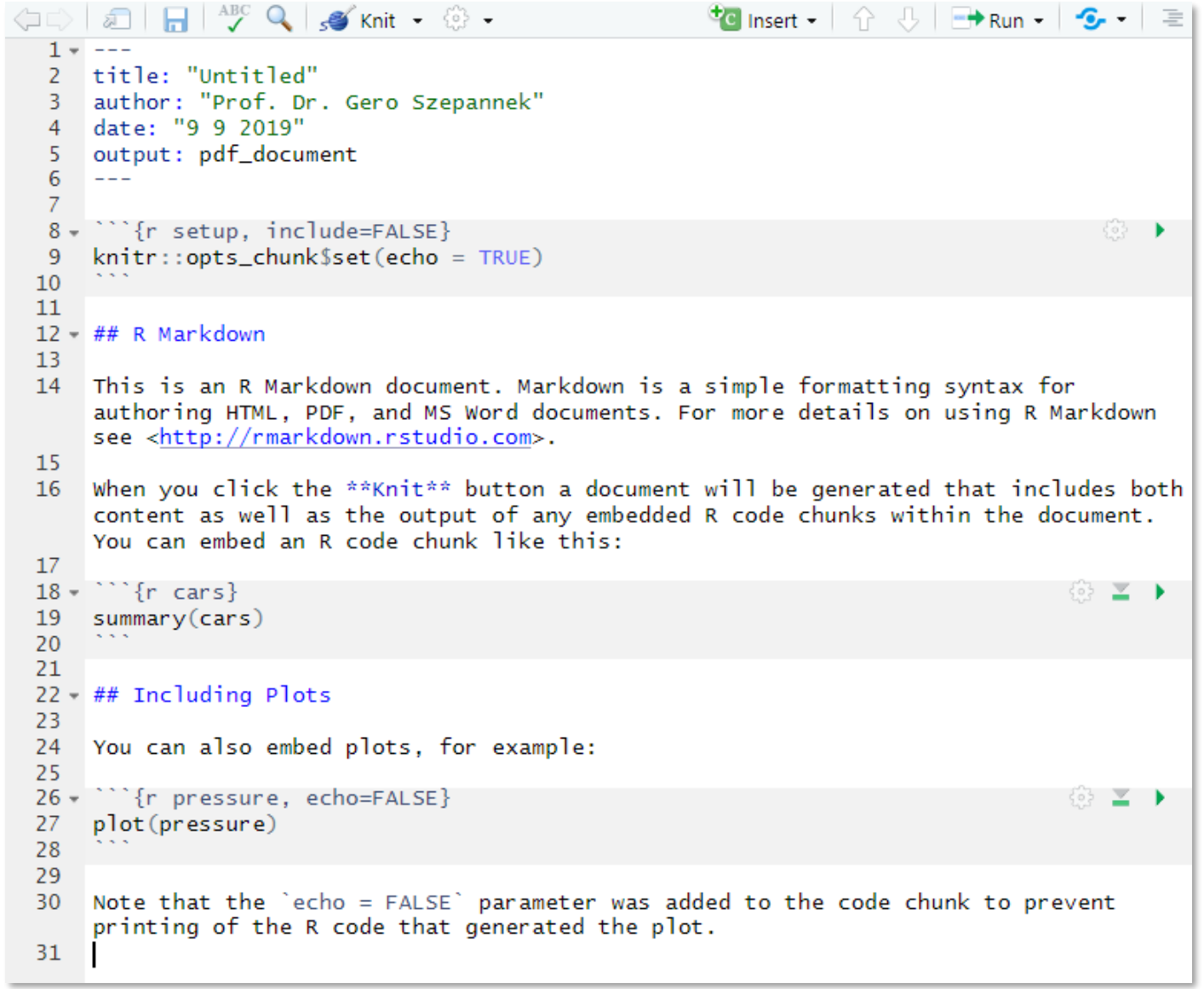

Gero Szepannek

# Literate Programming

“Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.” (D. Knuth (1984): Literate Programming, The Computer Journal 27 p. 97.)


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
New File
New Project...
Open File... Ctrl+O
R Script Ctrl+Shift+N
R Notebook
R Markdown...


New R Markdown
Document
Presentation
Shiny
From Template
Title: Untitled
Author: Prof. Dr. Gero Szepannek
Default Output Format:
HTML
Recommended format for authoring (you can switch to PDF or Word output anytime).
PDF
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).
Word
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).
OK
Cancel


Knit
Insert
Run
---
title: "Untitled"
author: "Prof. Dr. Gero Szepannek"
date: "9 9 2019"
output: pdf_document
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
## R Markdown
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
```{r cars}
summary(cars)
```
## Including Plots
You can also embed plots, for example:
```{r pressure, echo=FALSE}
plot(pressure)
```
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Knit Insert Run

```
---
title: "Untitled"
author: "Prof. Dr. Gero Szepannek"
date: "9 9 2019"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

## Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

Meta-information

(Sub-)Section header

Link

R chunk

R chunk where code is not shown in output but only the figure

****bold****
**italic**

https://rmarkdown.rstudio.com/

https://bookdown.org/yihui/rmarkdown/

# Exercise

1. Create a Rmarkdown template as shown on the previous slides!
2. Run the code find out what documents are created!
3. Modify the template such that a boxplot of the variable age of the custdata.csv is displayed.

# Exercise / Case Study: Identify Customers who don‘t have a health insurance

**Assignment of teams:**

1. Import the file custdata.csv!
2. Create a Rmarkdown file to present your results!
3. How many observations have the data?
4. What variables are in the data? Which variable is the target variable that indicates whether a customer has a health insurance?
5. What percentage of customers has no health insurance?
6. What can you tell about missing data (NA)? (Note: the function `is.na()` checks for missings.)
7. What can you tell about the variables income and age with regard to plausibility?
8. Create a table to compare whether the percentages of insured persons varies by the state (state.of.res). Do also consider absolute numbers. What would you conclude?
9. Create some boxplots and mosaicplots: which variable can be used to identify persons that don‘t have a health insurance?
10. **Submit your team‘s results a a .Rmd that should run on my computer.**