Introduction to R for data analysis

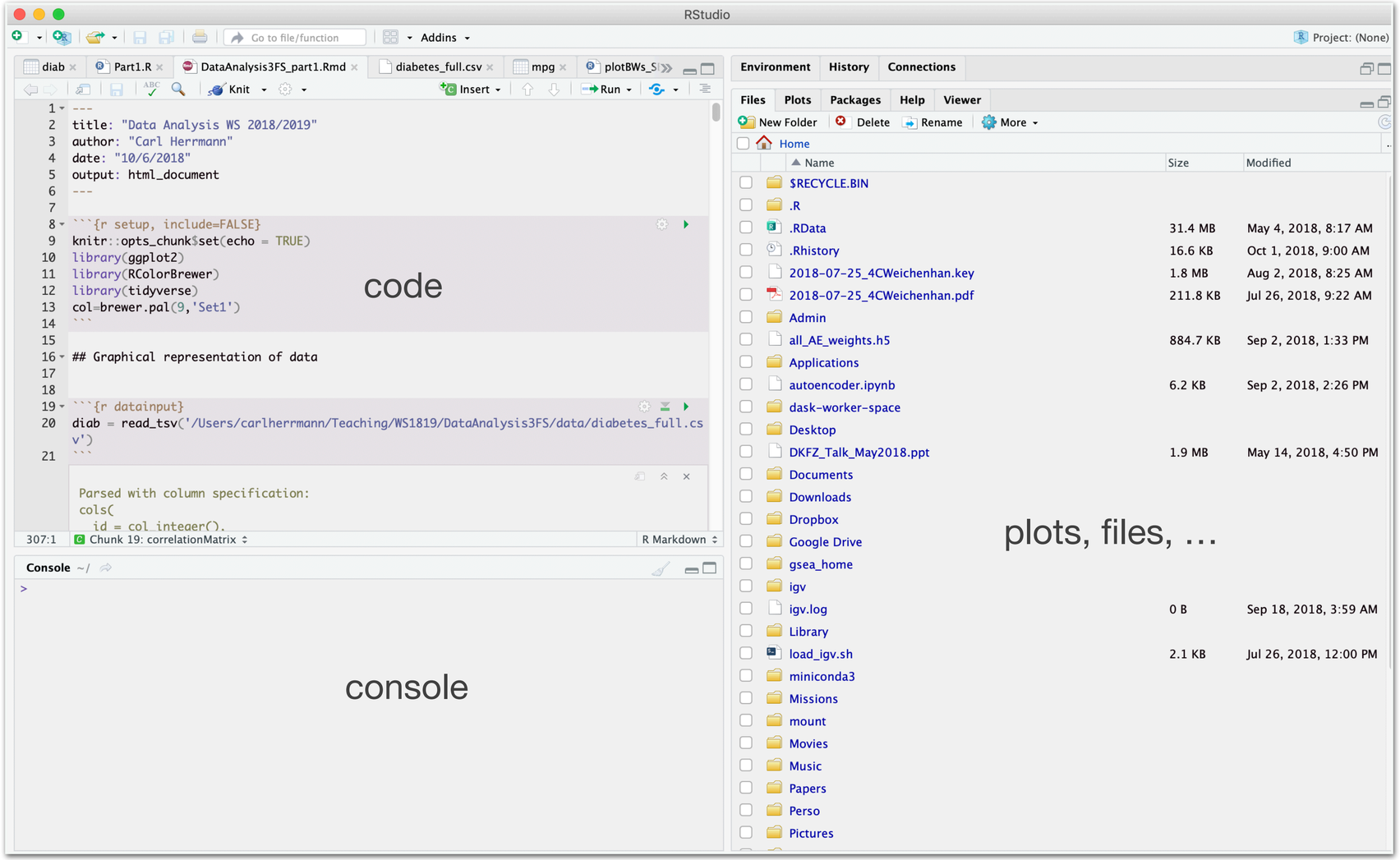
1. RStudio / Rmarkdown

Carl Herrmann & Carlos Ramirez
R4SC - Freiburg June 2024

IPMB
Institut für Pharmazie und Molekulare Biotechnologie

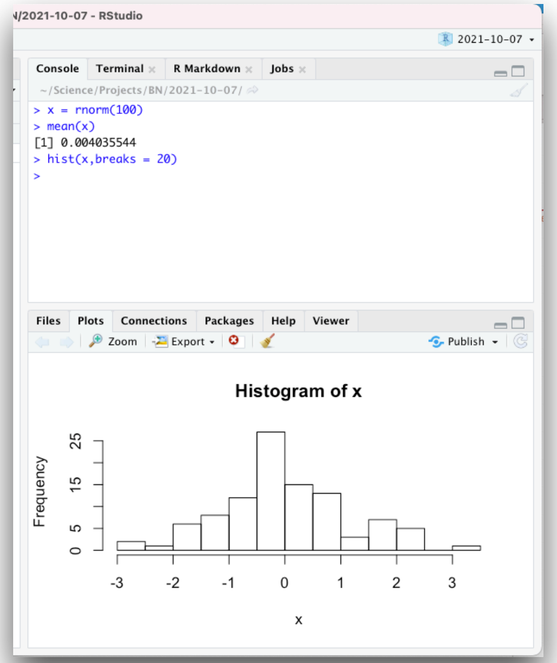
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

RStudio

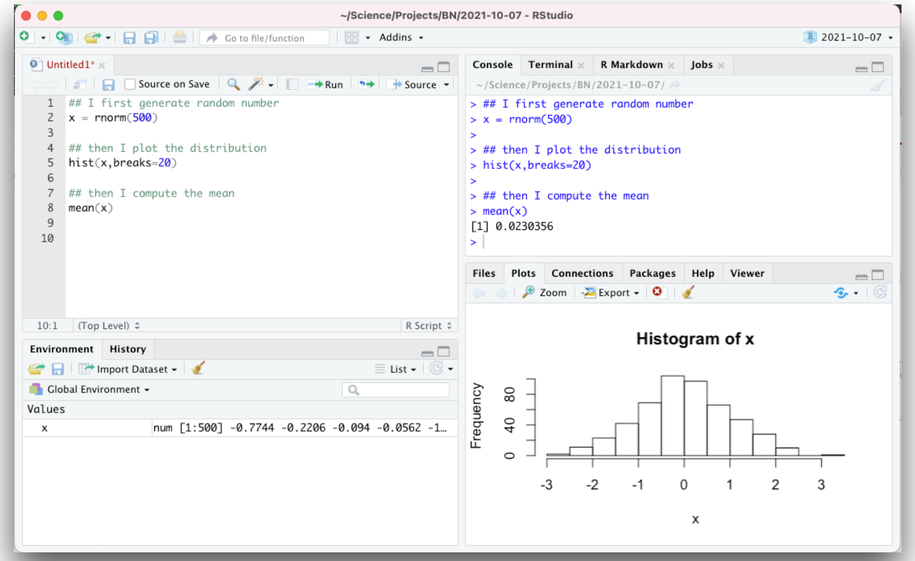

RStudio
Go to file/function
Addins
Project: (None)
diab
Part1.R
DataAnalysis3FS_part1.Rmd
diabetes_full.csv
mpg
plotBWs_S
Knit
Insert
Run
1 ---
2 title: "Data Analysis WS 2018/2019"
3 author: "Carl Herrmann"
4 date: "10/6/2018"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 library(ggplot2)
11 library(RColorBrewer)
12 library(tidyverse)
13 col=brewer.pal(9,'Set1')
14 ```
15
16 ## Graphical representation of data
17
18
19 ```{r datainput}
20 diab = read_tsv('/Users/carlherrmann/Teaching/WS1819/DataAnalysis3FS/data/diabetes_full.csv')
21 ```
Parsed with column specification:
cols(
id = col_integer(),
code
307:1
Chunk 19: correlationMatrix
R Markdown
Console ~/
>
console
Environment
History
Connections
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Name
Size
Modified
$RECYCLE.BIN
.R
.RData 31.4 MB May 4, 2018, 8:17 AM
.Rhistory 16.6 KB Oct 1, 2018, 9:00 AM
2018-07-25_4CWeichenhan.key 1.8 MB Aug 2, 2018, 8:25 AM
2018-07-25_4CWeichenhan.pdf 211.8 KB Jul 26, 2018, 9:22 AM
Admin
all_AE_weights.h5 884.7 KB Sep 2, 2018, 1:33 PM
Applications
autoencoder.ipynb 6.2 KB Sep 2, 2018, 2:26 PM
dask-worker-space
Desktop
DKFZ_Talk_May2018.ppt 1.9 MB May 14, 2018, 4:50 PM
Documents
Downloads
Dropbox
Google Drive
gsea_home
igv
igv.log 0 B Sep 18, 2018, 3:59 AM
Library
load_igv.sh 2.1 KB Jul 26, 2018, 12:00 PM
miniconda3
Missions
mount
Movies
Music
Papers
Perso
Pictures
plots, files, …

Using RStudio

- **Option 1** : execute R code in the console
  - cannot be easily reproduced
  - gone if you exit RStudio

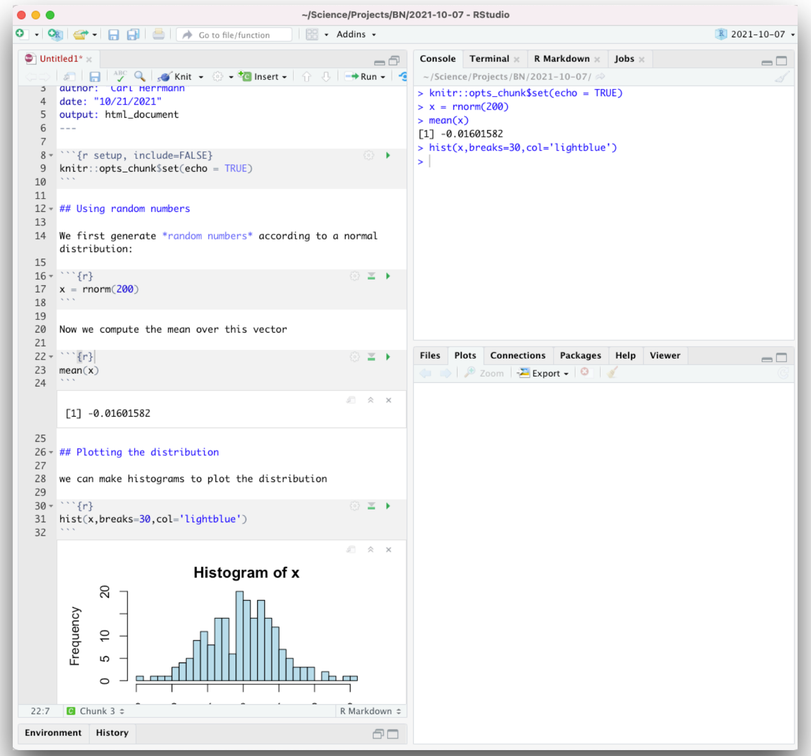


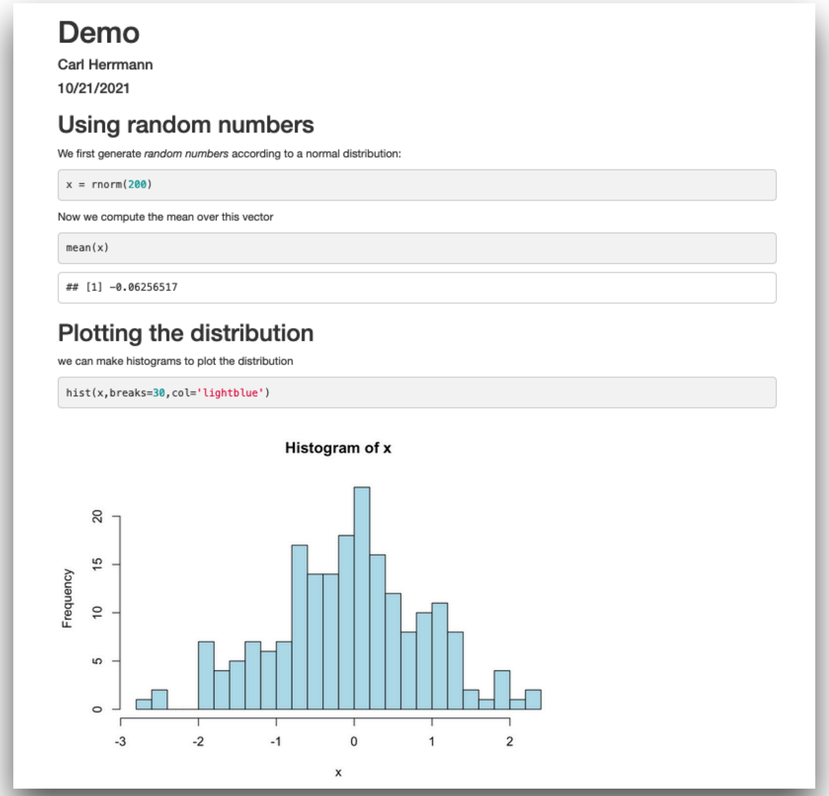
- **Option 2** : write an R script
  - you can save the script and execute it again



Using RStudio

- **Option 3** : write an Rmarkdown document
  - creates nice pdf / html reports
  - dynamic: include R code that is executed evry time you change something



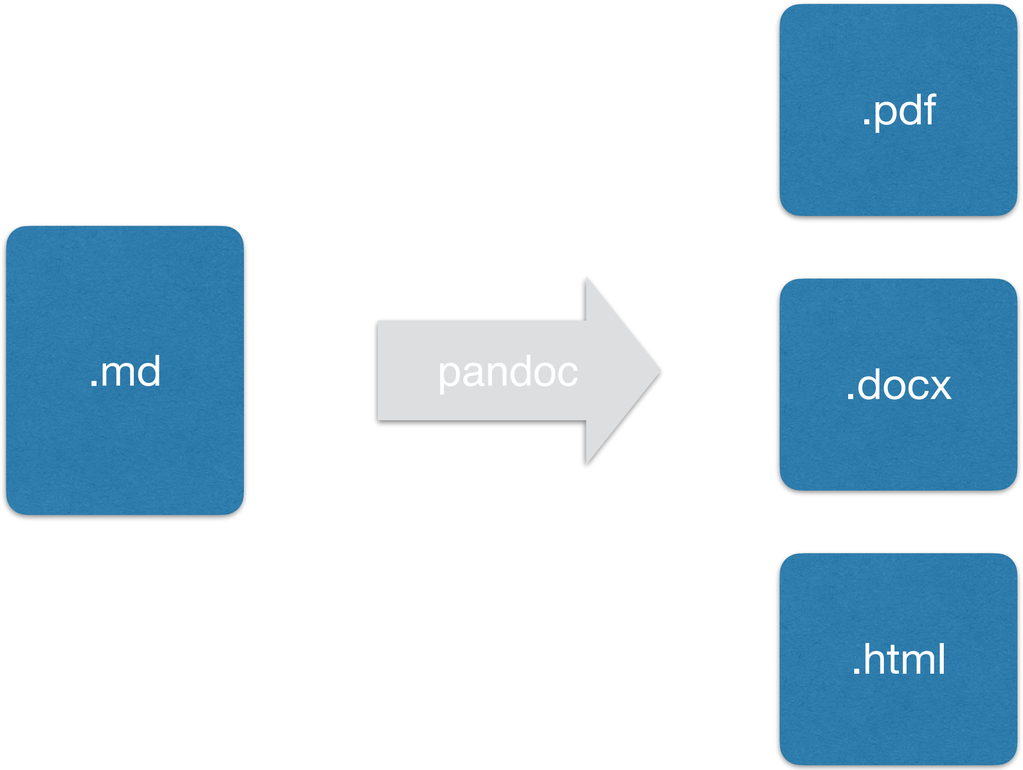


Reproducible analysis

- **Reproducibility** of results is important for experimental work (lab book!) but also for computational analysis!
- Different systems have been developed to ensure reproducibility of analysis flows
- For R based analysis, the best option is to use **Rmarkdown** documents
  - high levels of reproducibility
  - easy update of results
  - easy sharing of results with colleagues (html / pdf document)

Markdown

- Markdown is a way to format plain text with a simple text editor
- Markdown documents can be converted with a **renderer** into
  - html
  - pdf
  - word



Rendering markdown

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

markdown

```
# My document

## this is a header

In the text we can *highlight* or put in **bold**.

## making lists

We can make **numbered lists**:

1. first item
2. second item

or unordered lists

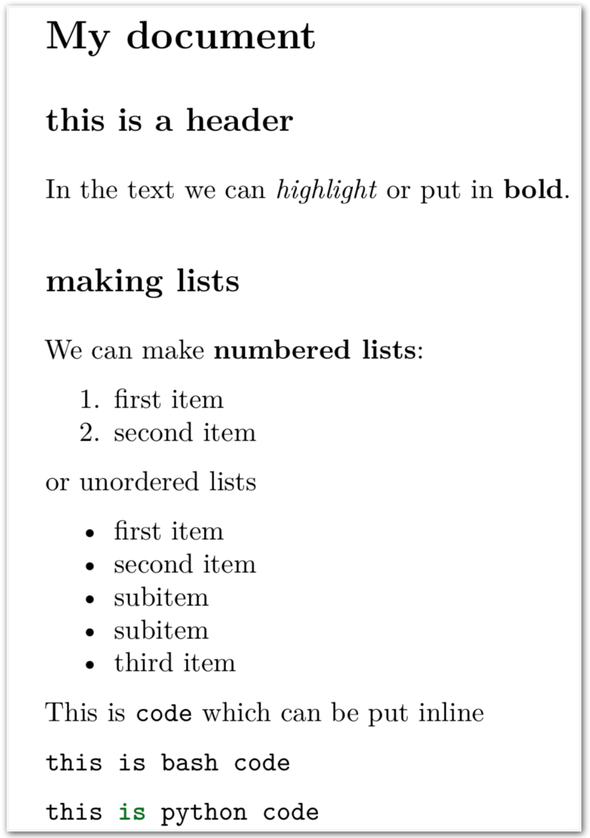
* first item
* second item
  + subitem
  + subitem
* third item

This is `code` which can be put inline

```bash
this is bash code
```

```python
this is python code
```
```

pdf



My document

this is a header

In the text we can *highlight* or put in **bold**.

making lists

We can make **numbered lists**:

1. first item
2. second item

or unordered lists

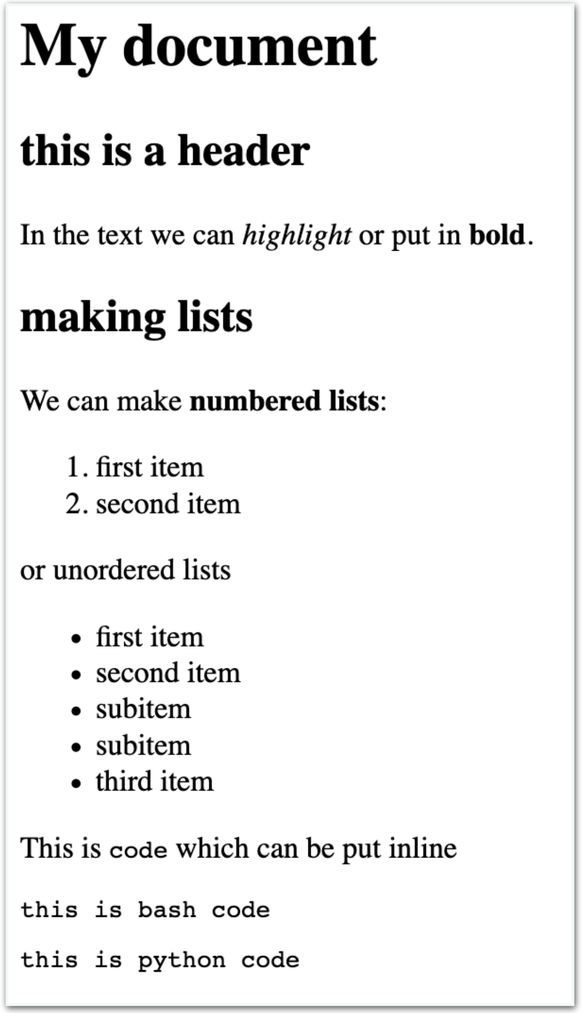
- first item
- second item
- subitem
- subitem
- third item

This is `code` which can be put inline

```
this is bash code
```

```
this is python code
```

html



My document

this is a header

In the text we can *highlight* or put in **bold**.

making lists

We can make **numbered lists**:

1. first item
2. second item

or unordered lists

- first item
- second item
- subitem
- subitem
- third item

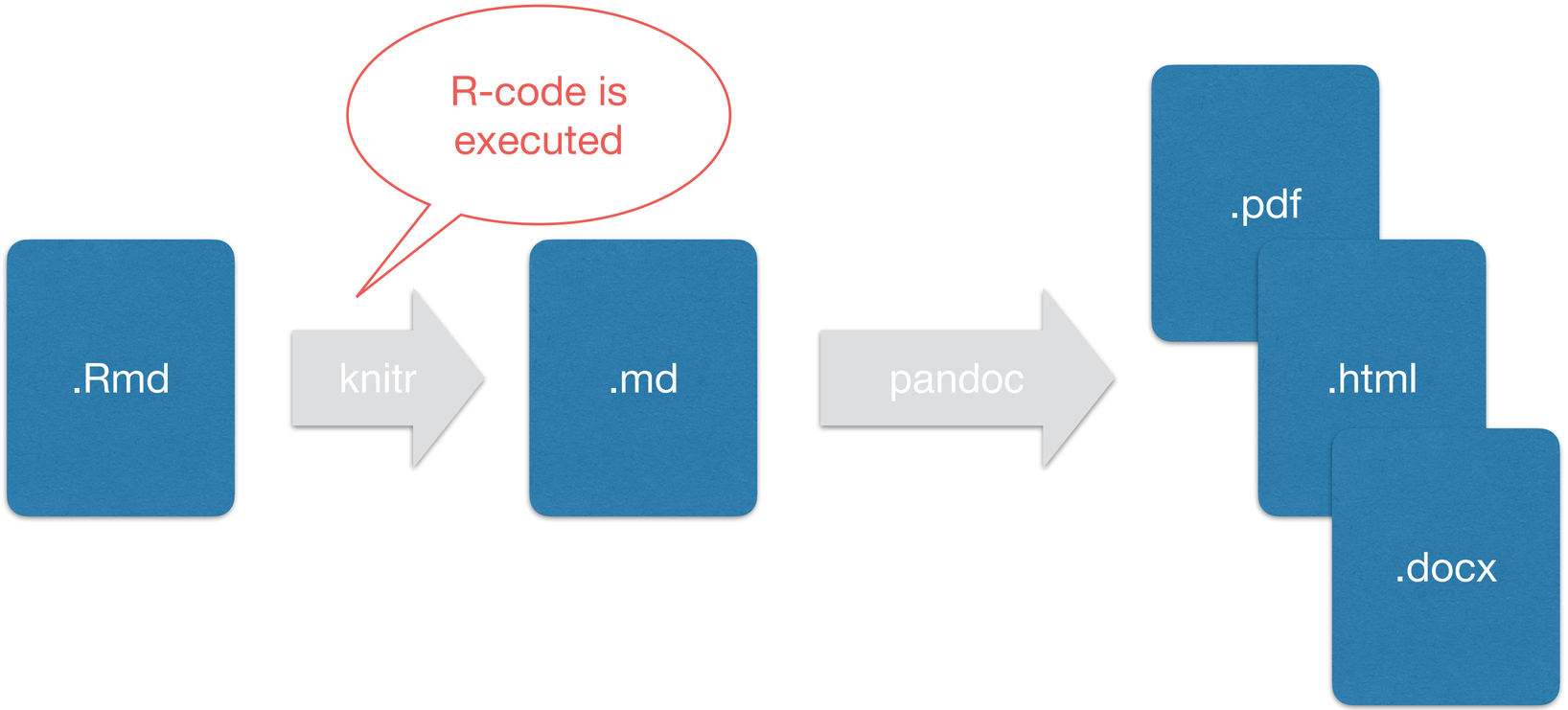
This is `code` which can be put inline

```
this is bash code
```

```
this is python code
```

Rmarkdown

- With Rmarkdown, R-code parts can be included into the markdown document
- the R-code will be executed, the result integrated into markdown



Rmarkdown format

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Knit | Insert | Run

```
---
title: "Demo"
author: "Carl Herrmann"
date: "6/14/2021"
output: html_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, an
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 lik
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.
```

Demo

Carl Herrmann

6/14/2021

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.

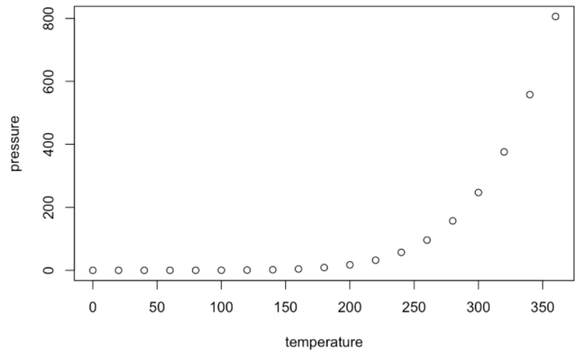
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



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

Rmarkdown format

IPMB
Institut für Pharmazie und Molekulare Biotechnologie
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

````
---
title: "Project 01"
author: "Carl Herrmann"
date: "4/17/2019"
output:
  html_document:
    keep_md: yes
  pdf_document: default
---
# A Rmarkdown tutorial

This is a brief tutorial on how to use Rmarkdown to create dynamic documents

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_knit$set(root.dir='/Users/carlherrmann/Teaching/SS2019/DataAnalysis_4FS')
```

## Load the dataset

```{r read_data}
allDepMapData = readRDS('Data/depmap/DepMap19Q1_allData.RDS')
```

Now plot the distribution of the cell lines according to the tissue type

```{r plot_data}
freq = sort(table(allDepMapData$annotation$Primary.Disease))
par(las=2,mar=c(3,8,3,3));barplot(freq,horiz=TRUE, col='lightgrey')
```
````

header: set options

R code chunks

text in markdown

Rmarkdown chunk options

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

valid for all chunks

- echo=TRUE : R-code is displayed in final document
- cache = TRUE : results of all chunks are cached

- Display options can be set for each chunk individually, or for all chunks at the beginning of the document

```
```{r plot_data,fig.height=12,fig.width=12}
freq = sort(table(allDepMapData$annotation$Primary.Disease))
par(las=2,mar=c(3,8,3,3));barplot(freq,horiz=FALSE, col='lightgrey')
```
```

valid for **this** chunks

- set height and width of output figure
```

Reference

- https://rmarkdown.rstudio.com/
- https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf