```r
library(dplyr)

rladies_global %>%
  filter(city == 'Munich')
```


REPRODUCIBLE RESEARCH

With Rmarkdown; Git+GitHub+RStudio

1.
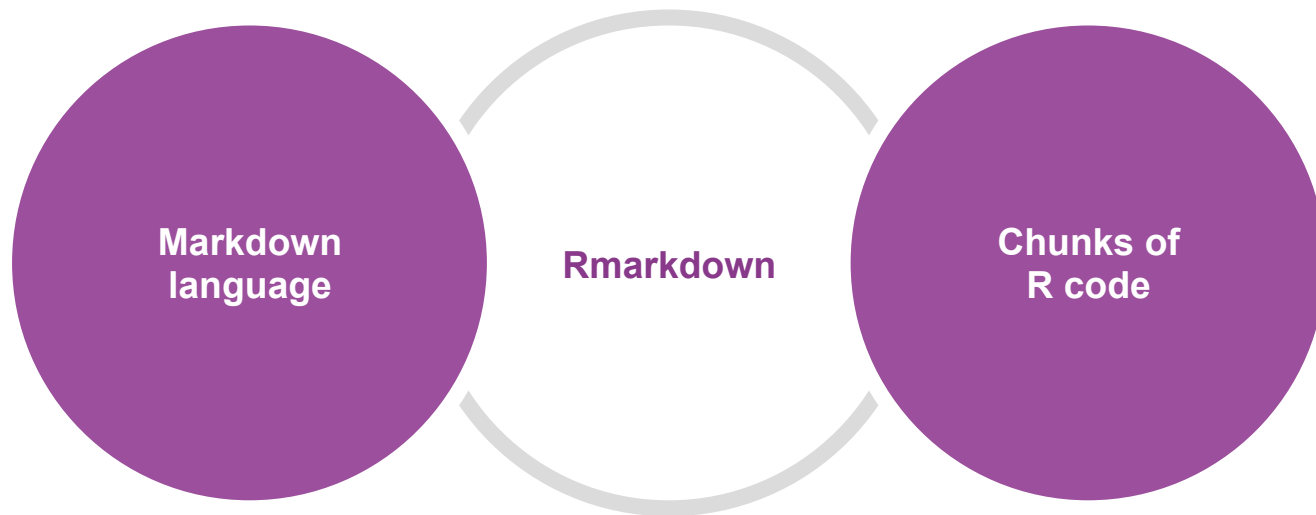# Why reproducible research?

Reproducible Research
Why?

- To improve trust in science

- Transparency: to allow others to verify our work

- To keep track of our procedure as future reference (especially for ourselves)

- To have a local or online backup, sorted by history.

2.
Rmarkdown

Rmarkdown
what is it?

Markdown language

Rmarkdown

Chunks of R code

Rmarkdown
what is it?

```
         My Report

This is the text of
my report. I have
some words and data
here.

I'd also like to
add a plot:

```{r}
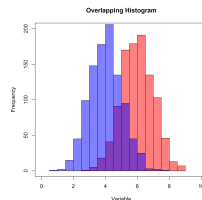Hist(data$myhist)
```
```

My Report

This is the text of my report. I have some words and data here.

I'd also like to add a plot:

- Html 5

- Pdf (Latex)

- Word document

- Slides

Rmarkdown
Why should I use it?

- It's easy to write and easy to read (even for humans)

- Different outputs: html, pdf, word, slides, etc.

- Includes both text and R code

- It is possible to apply LaTeX templates

Examples
from real life

`attach()`. I will be using `nhanesdataset_a.tsv` for the examples, but you may use whichever you like the most.

1. Using logical operators.

This is the way that Riccardo and Vindi showed on class, and the one that they recommend. The nice thing about this approach is that is very straightforward: if you understand the logic of "greater-than" and "smaller-than", is very easy to come up with this solution. Plus, you'd get to practice using R's subsetting and logical operators. From my point of view, the main disadvantage about this method is having to type too much.

```
age.cat1 <- NULL                          # new empty vector: age.cat1
age.cat1[age >= 20 & age <= 34] <- 1      # First level: 20-34
age.cat1[age >= 35 & age <= 49] <- 2      # Second level: 35-49
age.cat1[age >= 50 & age <= 64] <- 3      # Third level: 50-64
age.cat1[age >= 65 & age <= 79] <- 4      # Fourth level: 65-79
age.cat1[age >= 80] <-5                   # Fifth level: more than 80
age.cat1 <- as.factor(age.cat1)           # Convert age.cat1 into factor
tab$age.cat1 <- age.cat1                  # Add age.cat1 to dataset
summary(age.cat1)                         # check variable
```

```
##   1   2   3   4   5
## 761 800 786 556 206
```

As you can see, most of the subjects are included in levels 2, 3 and 1, respectively and in decreasing order. The group with less subjects is number 5.
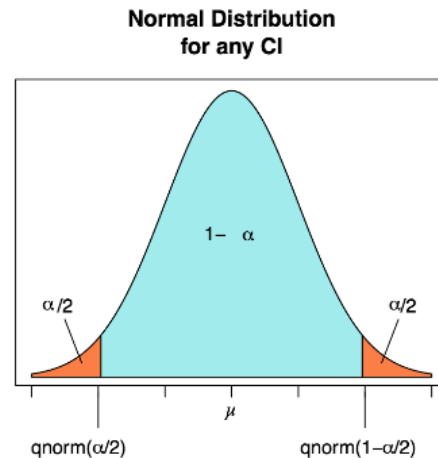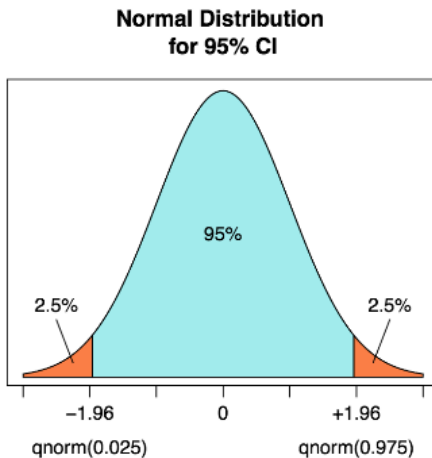
Examples
from real life

Manually estimating confidence intervals for a mean

Daloha Rodríguez-Molina

November 5, 2015

I'll try to replicate Riccardo's explanation about how to derivate the way we manually calculate confidence intervals in R, from the standard normal distribution.

All of the formulas we use from now on can be graphically understood by looking at these two graphs:

Normal Distribution for 95% CI

95%

2.5%   2.5%

−1.96   0   +1.96

qnorm(0.025)   qnorm(0.975)

Normal Distribution for any CI

$1 - \alpha$

$\alpha/2$ $\alpha/2$

μ

qnorm($\alpha/2$)   qnorm($1-\alpha/2$)

Rmarkdown
How do I do it?

```
# Header 1

### Header 3

**bold**

*italics*

inline equation: $A = \pi*r^{2}$

Image: ![](path/to/rladieslogo.png)

[link](www.rladies.org)
```

Header 1

Header 3

bold

Italics

inline equation: $A = \pi * r^2$

image:

[link]

3.
Next steps

What do I do
with my report?

- We want to keep track of our work

- We want to make it accessible for others to increase transparency

- We want others to collaborate, while we're able to keep track of changes.

- It'll probably change or improve after revisions

Proposed
solution
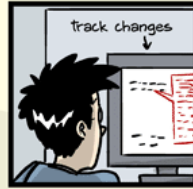

Curious? Check out: http://happygitwithr.com/

Now it's
YOUR turn!

- Convert the provided document into an Rmarkdown document following the indications.

- ## Link: http://bit.ly/2pltahB

- Use the Rmarkdown cheat sheet or ask me if you have any questions.

- If you want feedback, send your R code to daloha@rladies.org (optional)

THANKS!

Twitter: @darokun
    @RLadiesMunich

Email: daloha@rladies.org