World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

منظمة الصحة العالمية
المكتب الإقليمي لشرق المتوسط

REGIONAL TRAINING ON CAPACITY DEVELOPMENT OF DATA ANALYTICS AND DISSEMINATION USING "R" SOFTWARE

AMMAN, JORDAN, 3 - 7 DECEMBER, 2023

Day #4

HEALTH FOR ALL BY ALL
a call for solidarity and action

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

Outline

- Wrap-up
- Data visualization using ggplot2 packages
- Report generation with RMarkdown
- Q & A

Data visualization with ggplot2

Session 4 Agenda

- 9:00 – 9: 30 (30 min): **Wrap-up**
- 9:30 – 9:50 (20 min): **Presentation "Data visualization with ggplot2"**
- 9:50 – 10:20 (30 min): **Demonstration**
- 10:20 – 10:40 (20 min): **Stretching / coffee break**
- 10:40 – 12:30 (1.8 hr): **Practice/Exercise**
- 12:30 – 13:00 (30 min): **Quick debrief/ Q&A**
- 13:00 – 14:00 (60 min): **Lunch**
- 14:00 – 14:20 (20 min): presentation "Report generation with RMarkdown"
- 14:20 – 14:50 (30 min): Demonstration
- 14:50 – 15:10 (20 min): Stretching / coffee break
- 15:10 – 16:30 (80 min): Practice/Exercise
- 16:30 – 17:00 (30 min): Quick debrief/ Q&A

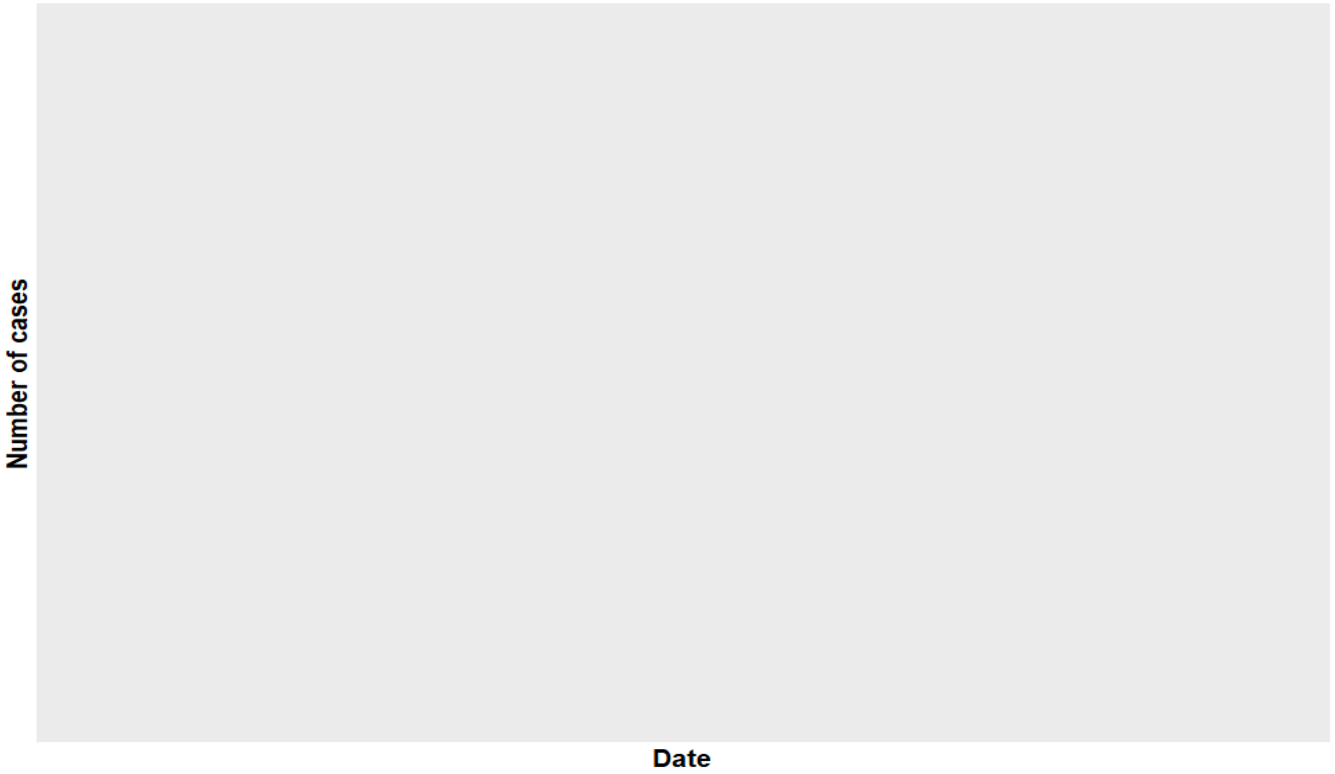
3. Time

Epidemic curve:

- **Components**

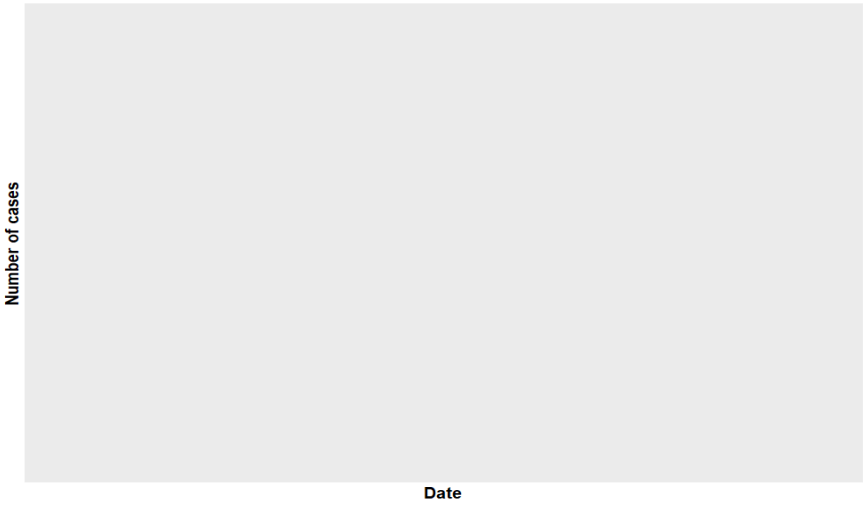
1. Number of cases (Y-axis)
2. Time (X-axis)
3. Grouping by another variable

(e.g. sex, country, hospital) [optional]



3. Time

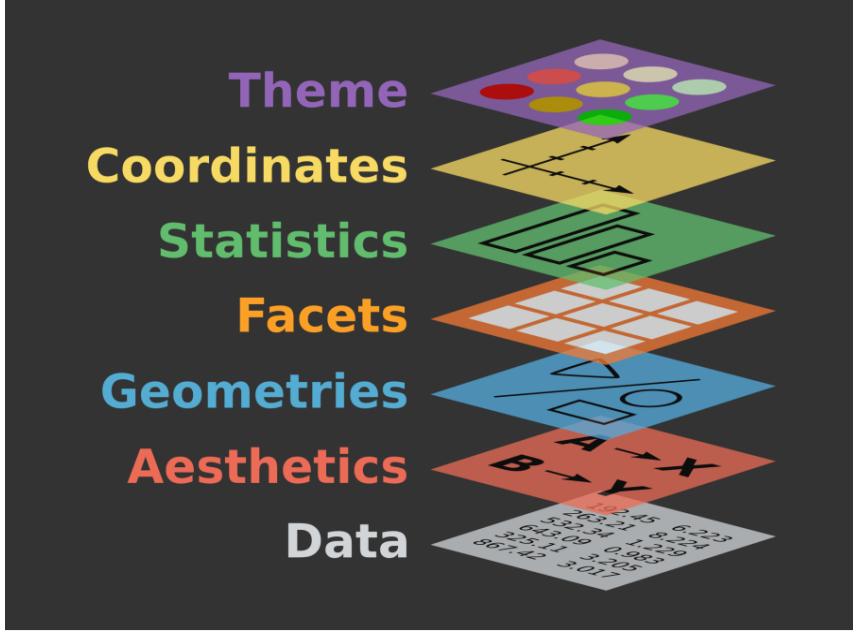
Epidemic curve:



- **Components**

1. Number of cases (Y-axis)
    1. Data from line list: each row is a case
    2. Aggregated data
2. Time to be displayed
    1. Which time? (date of admission, date of onset,..)
    2. Class is date (YYYY-MM-DD)
    3. Completeness of the date (NA and/or sequence)
    4. What is the time interval you want to display cases by? (for example, daily, weekly, biweekly, monthly)

3. Time



Epidemic curve:

➢ **Create epi curve**

- **First option:** ggplot2::ggplot()
  1. Data (your dataset) ------▸ inside the ggplot() +
  2. Aesthetics (X & Y) ------▸ need only to assign the date to the x-axis (in case of line list data) +
  3. Geometries and binwidth ------▸ geom_histogram() +
  4. Coordinated and themes

Next slide

Image source: adapted from datasciencebox.org

3. Time

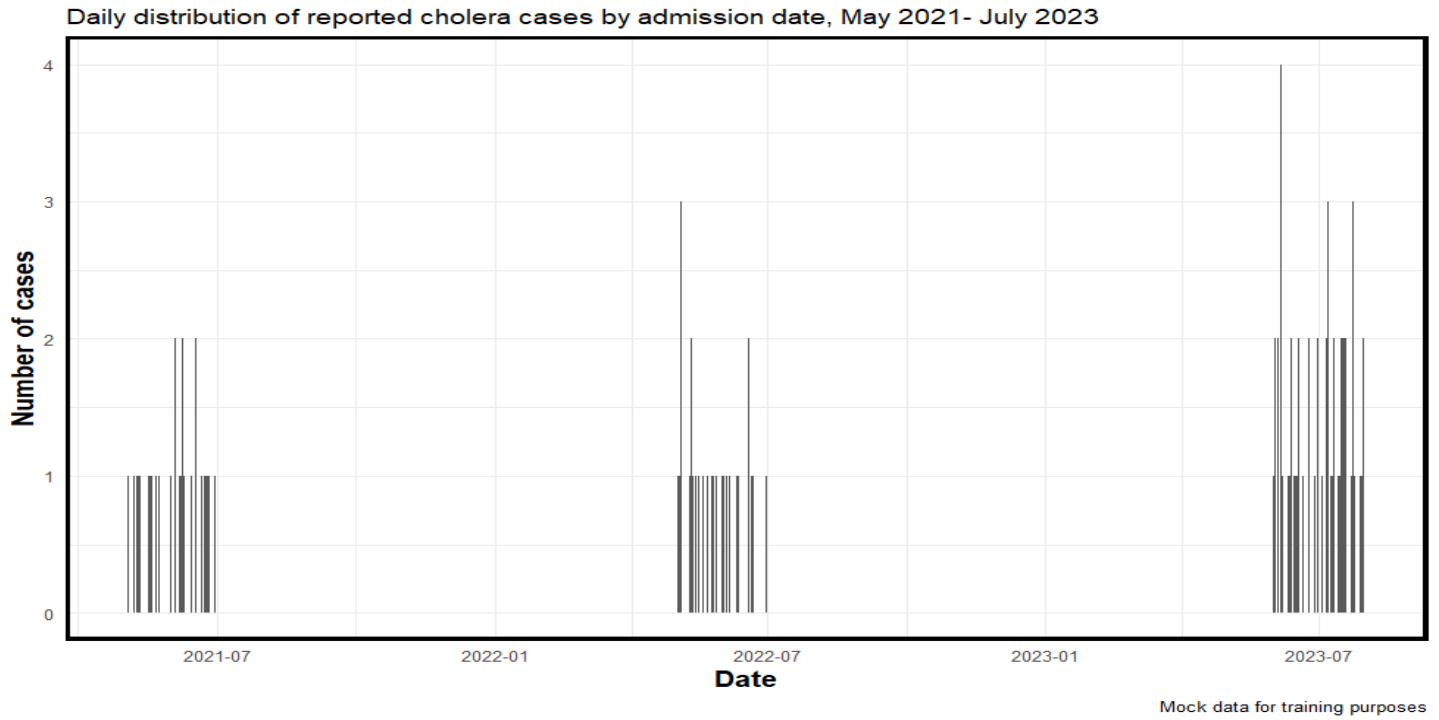
Epidemic curve:

First option:

ggplot2::ggplot()

Aggregate data to 1 day

```
cholera %>%  ggplot()+
  geom_histogram(aes(adm_date), binwidth = 1)+
  labs(title= "Daily distribution of reported cholera cases by admission date, May 2021- July 2023",
       x= "Date",
       y= "Number of cases",
       caption = "Mock data for training purposes")+
  theme_minimal()+
  theme(axis.title = element_text(face = "bold", size = 14),
        panel.border = element_rect(colour = "black", fill=NA, size=2))
```



3. Time

Epidemic curve:

First option: ggplot2::ggplot()

Aggregate data to week:
Here we need to create the weekly breaks beforehand to ensure that data (cases) are aggregated to the appropriate epi week with the following code

```
weekly_breaks <- seq.Date(
  from = floor_date(min(cholera$adm_date, na.rm = T), unit = "week", week_start = 7),
  to= ceiling_date(max(cholera$adm_date, na.rm = T), unit = "week", week_start = 7),
  by= "week"
)
```

Then, to use these weekly breaks as bins

A full code and epi cure in the next slide…

3. Time

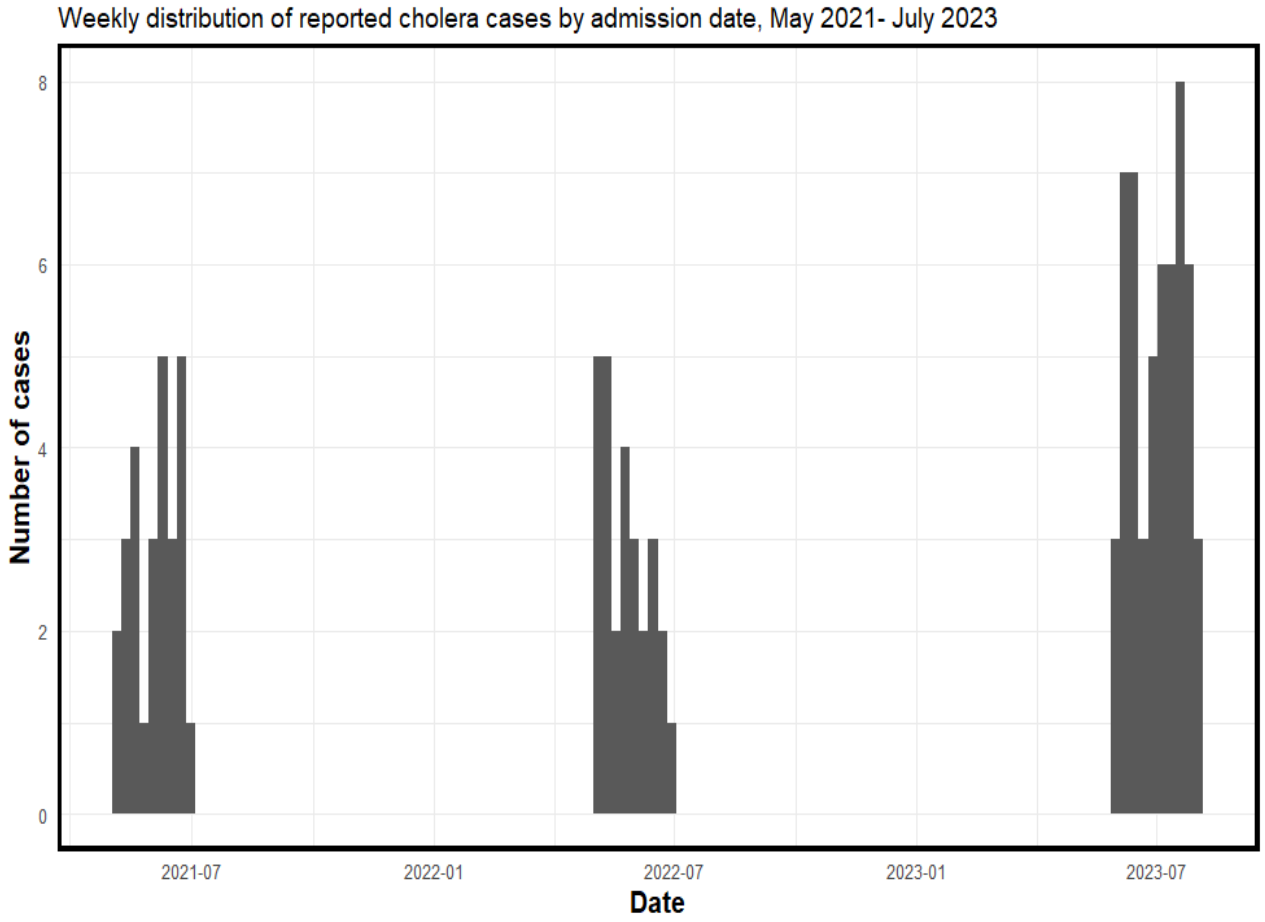
Epidemic curve:

First option: ggplot2::ggplot()

Aggregate data to week:

```
weekly_breaks <- seq.Date(
  from = floor_date(min(cholera$adm_date, na.rm = T), unit = "week", week_start = 7),
  to= ceiling_date(max(cholera$adm_date, na.rm = T), unit = "week", week_start = 7),
  by= "week"
)

cholera %>%
  ggplot(aes(adm_date))+
  geom_histogram(breaks= weekly_breaks,
                 closed= "left")
```



- Week_start= 7 >>> Sunday weeks
- Week_start= 1 >>> Monday weeks

3. Time

Epidemic curve:

- ➢ **Create epi curve**

- **Second option:** incidence::incidence()
  1. Data$date (your dataset)------▸ inside the incidence()
  2. Interval (to aggregate data to) ------▸ inside the incidence()

  +

  3. Plot ()

Next slide

HEALTH FOR ALL BY ALL a call for solidarity and action

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

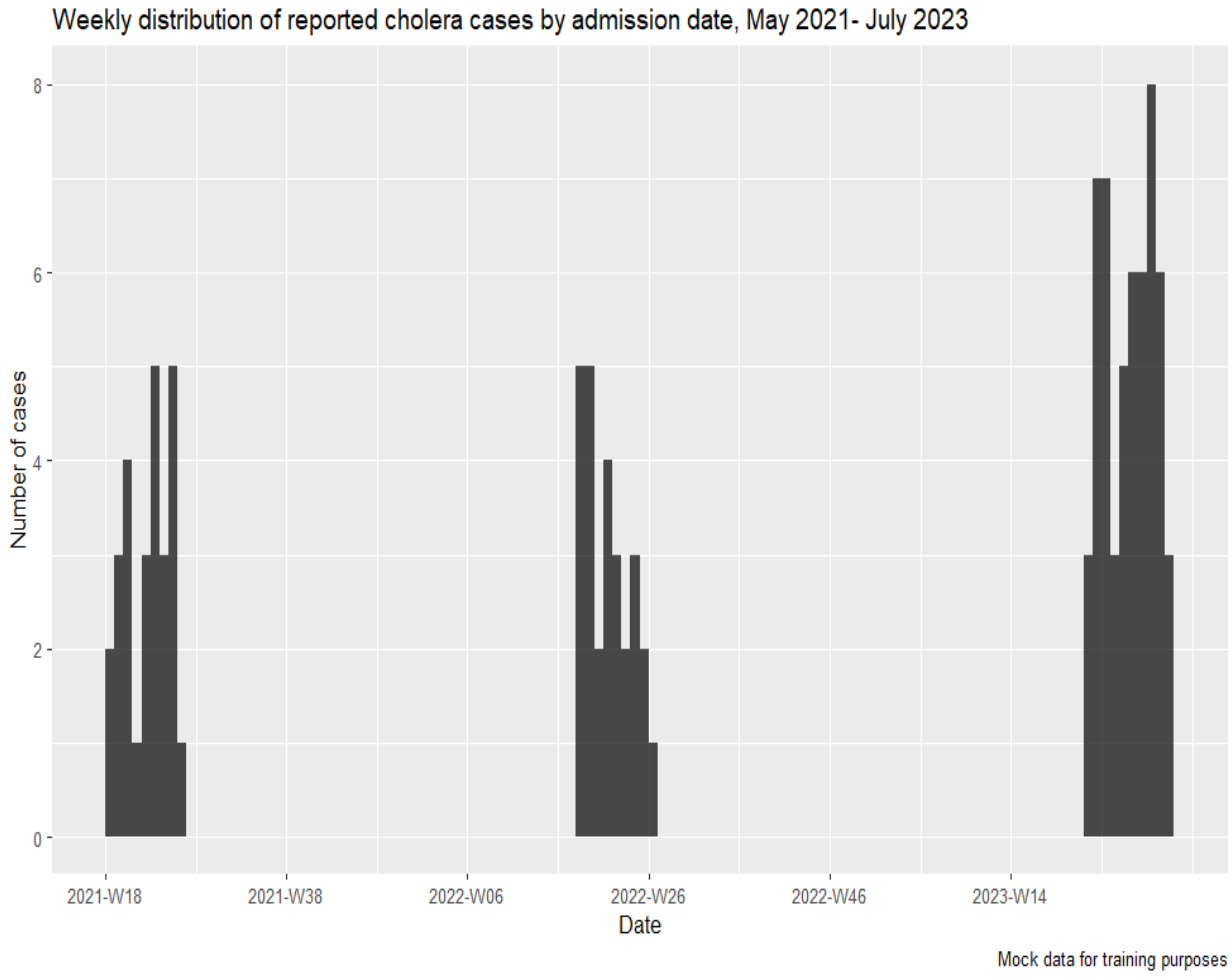
3. Time

Epidemic curve:

Second option: Incidence package

```
weekly_epi <- incidence(dates = cholera$adm_date, interval = "Sunday week")

plot(weekly_epi) +
  labs(title= "Weekly distribution of reported cholera cases by admission date, May 2021- July 2023",
       x= "Date",
       y= "Number of cases",
       caption = "Mock data for training purposes")
```



Exercise

- Open your training R project
- Create a new section “Time”
- Create an epi curve of daily distribution of reported cases by admission date
- Create an epi curve of the weekly distribution of reported cases by admission week
- Options:
  - ggplot2::ggplot()
  - incidence::incidence()

Report generation with RMarkdown

Session 4 Agenda

- 9:00 – 9: 30 (30 min): Wrap-up
- 9:30 – 9:50 (20 min): Presentation “Data visualization with ggplot2”
- 9:50 – 10:20 (30 min): Demonstration
- 10:20 – 10:40 (20 min): Stretching / coffee break
- 10:40 – 12:30 (1.8 hr): Practice/Exercise
- 12:30 – 13:00 (30 min): Quick debrief/ Q&A
- 13:00 – 14:00 (60 min): Lunch
- 14:00 – 14:20 (20 min): **presentation “Report generation with RMarkdown”**
- 14:20 – 14:50 (30 min): **Demonstration**
- 14:50 – 15:10 (20 min): **Stretching / coffee break**
- 15:10 – 16:30 (80 min): **Practice/Exercise**
- 16:30 – 17:00 (30 min): **Quick debrief/ Q&A**

RMarkdown

- It combines text (narrative) and R code outputs into an output document (word, pdf, HTML, PowerPoint)
- It creates automated, reproducible reports
- **It works with 2 main packages;**
  - **{rmarkdown}:** render the .Rmd into output
  - **{knitr}:** knit the code chunks into the document
  - **Pandoc:** software to convert the output into desired file type (*already installed with Rstudio*)

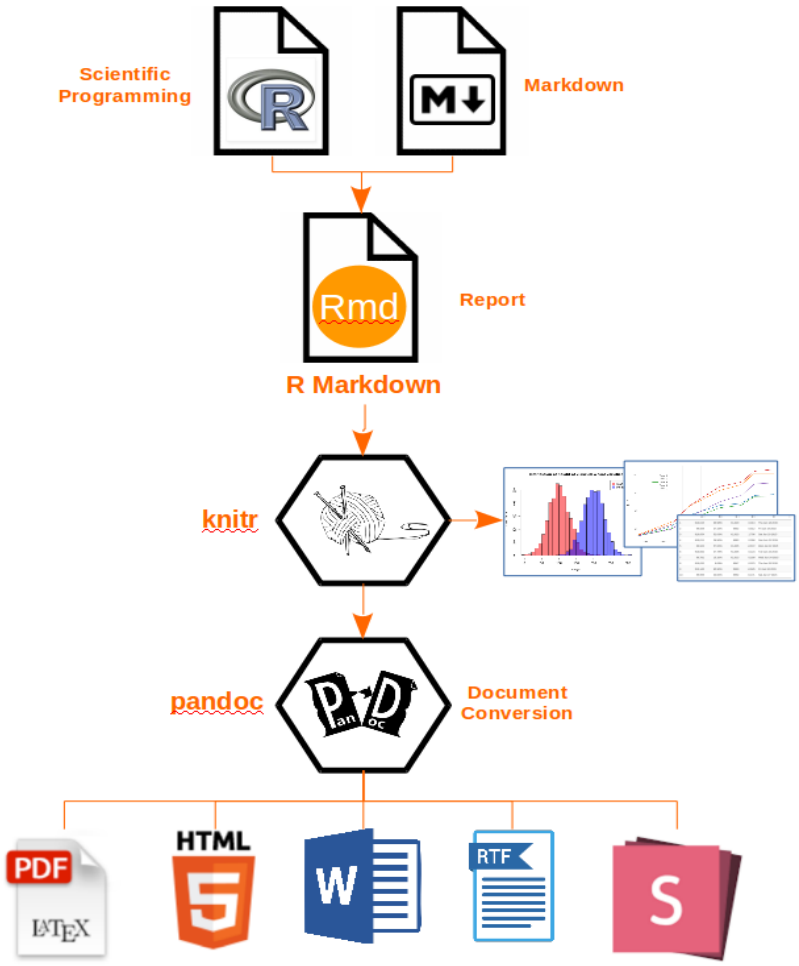
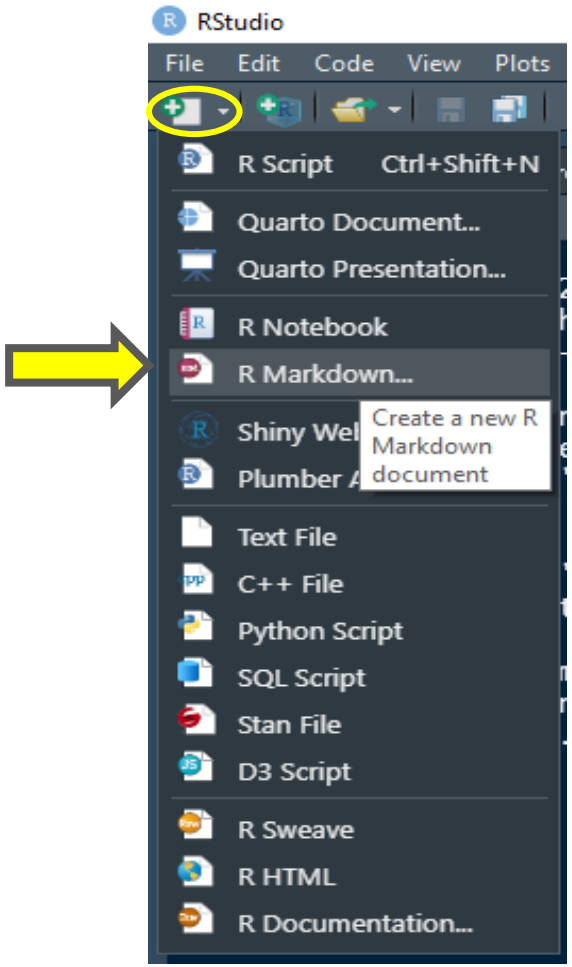
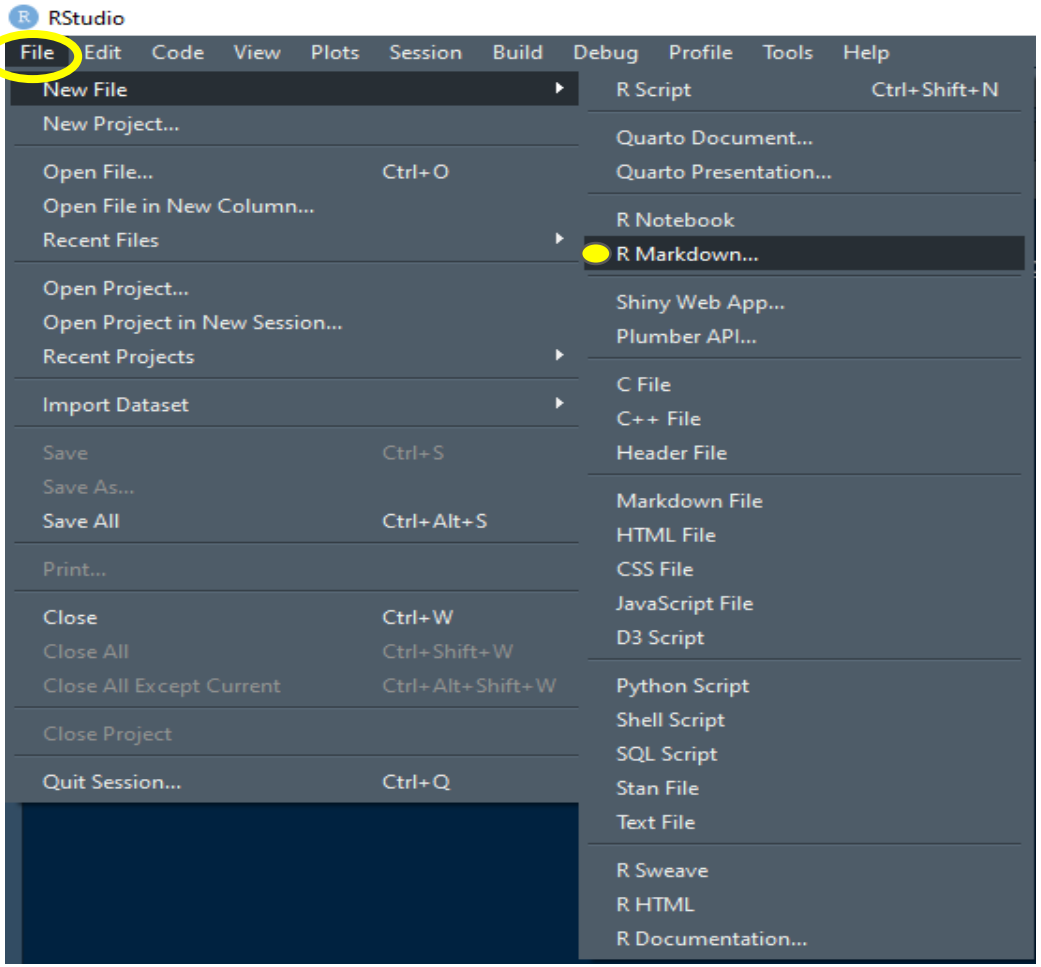


Image source: Make your research reproducible

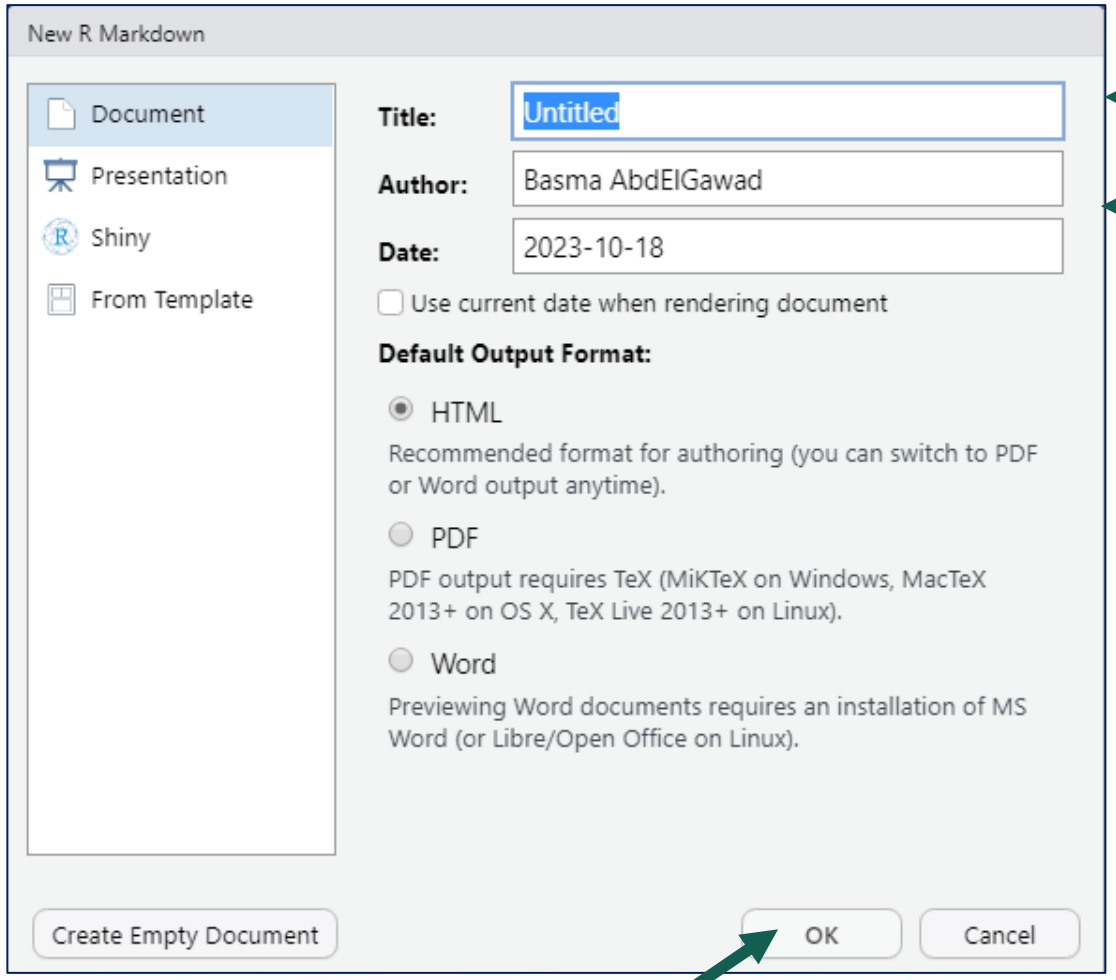
Create new RMarkdown report



OR



Create new RMarkdown report



Title of the report

Date of issuance, can be edited/ linked to current date (check the box)

Format of the report document

Click OK to create the RMarkdown file (next slide)

HEALTH FOR ALL BY ALL a call for solidarity and action

World Health Organization REGIONAL OFFICE FOR THE Eastern Mediterranean

RMarkdown

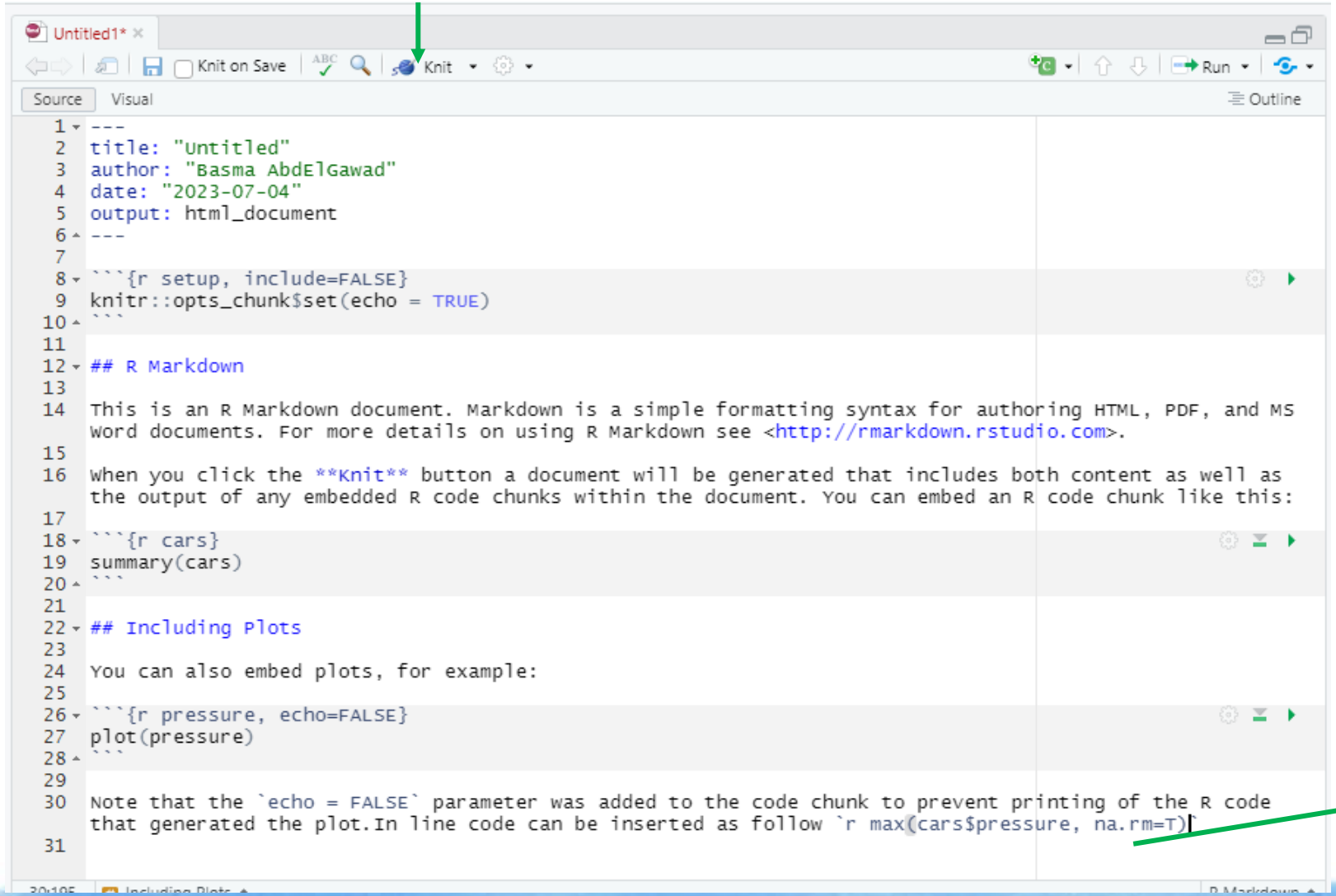
Produce the output document (final report)

YAML metadata

R-code chunk

Report Content (markdown text)

In-line code



Knitting to get output document report

demo_markdown.Rmd

Knit on Save | Knit | Run

Source | Visual | Outline

```
---
title: "Sample document output"
output: word_document
date: "2023-10-18"
---

```{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

Look at the 'render' pane in the console!

html

AutoSave Off

File | Home

Sample document output

2023-10-18

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.

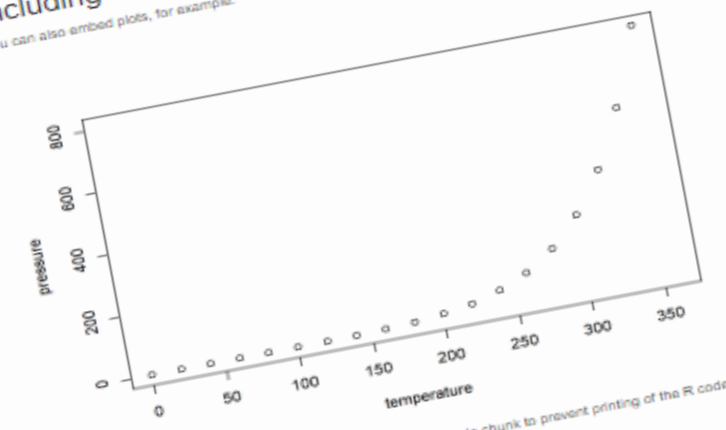
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

HEALTH FOR ALL BY ALL — a call for solidarity and action

World Health Organization — REGIONAL OFFICE FOR THE Eastern Mediterranean

Specific RMarkdown terminologies/shortcut

Term/shortcut	Meaning
• YAML	The top part of the document, start & end with - - -
• Inline code	Embed codes in markdown text, shown as `r `
• Code chunk	Includes R code sections, shown as ``` {r } ```
○ include = TRUE	Print R output to document
○ eval = TRUE	Run/evaluate R code
○ echo = TRUE	Print R code to the document
○ warning = TRUE	Print warnings to the document
○ message = TRUE	Print messages to the document
Ctrl + Alt + i	Create a new code chunk
Ctrl + shift + c	Comment/inactivate the markdown text

Narrative formatting in RMarkdown

Syntax	Becomes
`Plain text`	Plain text
`End a line with two spaces` `to start a new paragraph.`	End a line with two spaces to start a new paragraph.
`*italics* and _italics_`	*italics* and *italics*
`**bold** and __bold__`	**bold** and **bold**
`superscript^2^`	superscript^{2}
`~~strikethrough~~`	~~strikethrough~~
`[link](www.rstudio.com)`	link
`# Header 1`	**Header 1**
`## Header 2`	**Header 2**
`### Header 3`	**Header 3**
`#### Header 4`	**Header 4**
`##### Header 5`	**Header 5**
`###### Header 6`	**Header 6**

Image source: RMarkdown reference guide

HEALTH FOR ALL BY ALL
a call for solidarity and action

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean

Demonstration

Exercise

- Open your training R project
- Create a new RMarkdown script
- In the popped window:
  - Write the title: “Cholera cases till 2023-07-31”
  - Choose type of output document (this could be changed later)
- Save the RMarkdown script in the “script” folder in your project
- Explore tabs in the top bar of the script
- Knit your RMarkdown document (so far, rendered document will contain default RMarkdown script)
- Where is the output document located?
- Start editing the YAML section to be similar to this

```yaml
---
title: "Cholera cases till 2023-07-31"
author: "write your name and remeber you may remove it completely"
output: word_document
date: "`r Sys.Date()`"
---
```

- Keep the universal code chunk and adjust its options (warning = FALSE, message= FALSE, echo= FALSE)
- Remove all other chunks and texts

you can knit your document as many times as you wish!

Exercise cont.

- Add your code chunks for the following steps (*you did over the few past days*);
    1. Load required packages (give chunks simple meaningful names(optional))
    2. Import dataset "cholera_20231102.csv"
    3. Data processing
    4. Table1 [**title:** Reported cholera cases by age categories, till 31 July 2023]
    5. Graph1 [**HINT:** code for the epi curve, **title:** Distribution of reported cholera cases by date of admission and case category, till 31 July 2023]
- Copy your codes from your script or rewrite them and add them to the relevant chunks
- Try to run each chunk (‼*where the output occurs?*)
- Write executive summary text after loading package, importing data, and data processing chunks that show:
    - Heading "Executive summary"
    - Text: As of 2023-07-31, there were a total of XXX confirmed cholera cases and XXX deaths (CFR = xx%). The mean age of confirmed cases was XX years (standard deviation = xx)
- Add a title for the produced table and figure
- Knit your document

HEALTH FOR ALL BY ALL
a call for solidarity and action

World Health Organization
REGIONAL OFFICE FOR THE Eastern Mediterranean