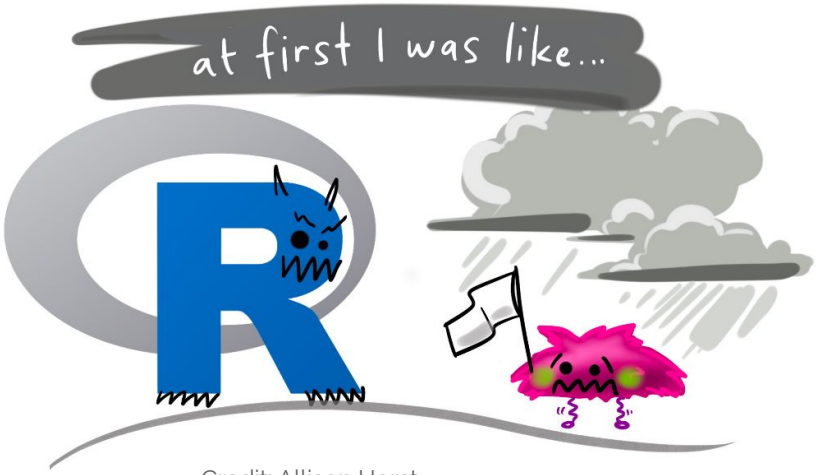
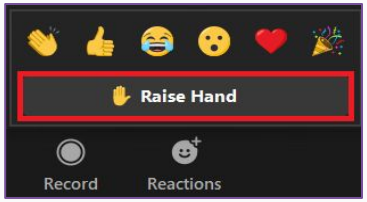
لغة الآر من الصفر إلى الإحتراف



Credit: Allison Horst

| **Twitter**:@batool664 | **Web**: https://batool-almarzouq.netlify.app/ | د.بتول المرزوق



قواعد الدورة

مرحبا في كورس "لغة الأر (R) من الصفر إلي الإحتراف"

المدربة: د. بتول المرزوق

الأسبوع: الأول

التاريخ: ٢٢ يونيو ٢٠٢٢

وصف الجلسة:

في هذه الجلسة سنتعرف على الخطة التعليمية مع بيان أهمية استخدام لغة آر (R) وإبراز دورها المميز في علم البيانات. لغة آر (R) هي لغة مفتوحة المصدر تُستخدم في التمثيل المرئي للبيانات وتعديلها، بالإضافة إلى النمذجة. وقد احتلت أعلى المراتب في تصنيفات لغات البرمجة المستخدمة في علوم البيانات على مدار عدة سنوات، مما أدى إلى تزايد استخدامها.

تسجيل الأسماء

الاسم / المدينة / حسابات وسائل التواصل الإجتماعي لمن يرغب (twitter و GitHub وما إلى ذلك)

-
-
-
-
-
-
-
-

- كتابة الاسم في Google Doc
  - الإجابة على Icebreaker question
- إطفاء الجوال
- يُفضل فتح الكاميرا
- لاتتردد في طرح **أي سؤال** خلال الجلسة
- **Slack**

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

مراجعة

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

Tidy Data

"TIDY DATA is a standard way of mapping the meaning of a dataset to its structure."

—HADLEY WICKHAM

كل متغير هو عمود

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

كل عينة هي صف

Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10


tidyverse
readr
ggplot2
forcats
tidyr
tibble
dplyr
stringr
purrr

كيفية قراءة البيانات في R

```
read_csv("path")
```

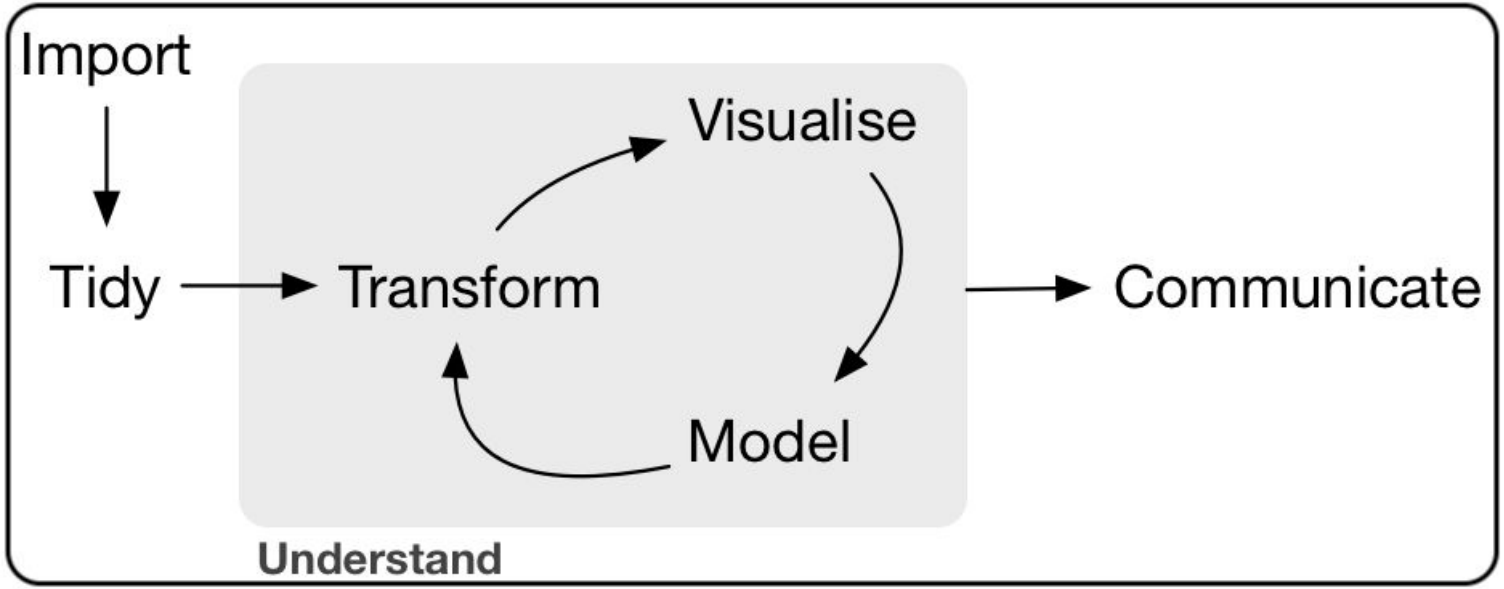
In relative to the working directory using `setwd()`

```
read_csv(here("folder","file"))
```

```
read_csv(here("data","data.csv"))
```

كيفية قراءة البيانات في R

Credit: Allison Horst


Import
Tidy
Transform
Visualise
Model
Understand
Communicate


DATAFRAME
$ OPERATOR
COLUMN TO SELECT
dataf$NewData
[1] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A"

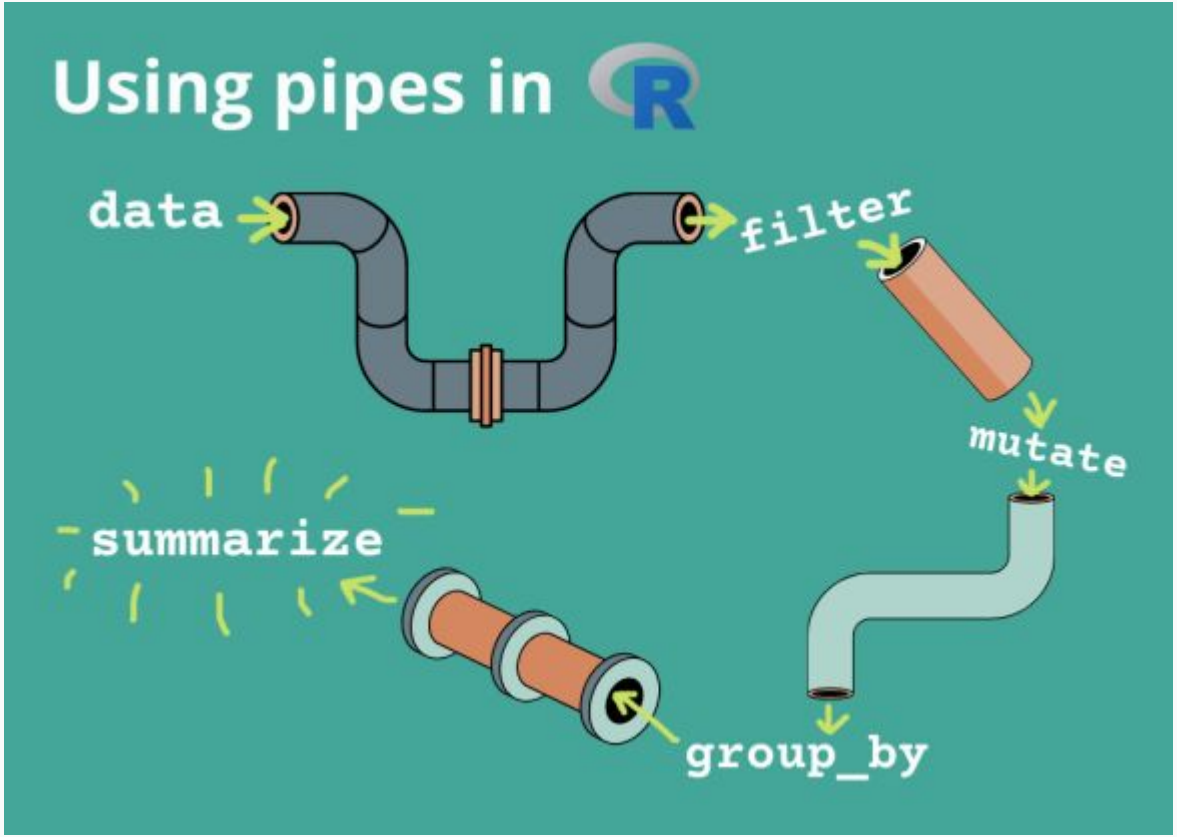
معنى هذا الرمز في لغة R

%>% = (و ثُم)

معنى هذا الرمز في لغة R

%>% = (و ثُم)

```
dataset %>%
select(column)
```



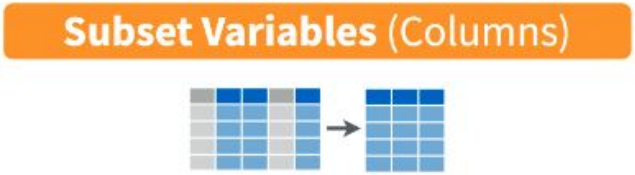
```
filter(data, row == "Mexico")
```

```
filter(data, row == "Mexico")
```

```
Data %>%
filter(row == "Mexico")
```

- `filter()` : pick observations by their values

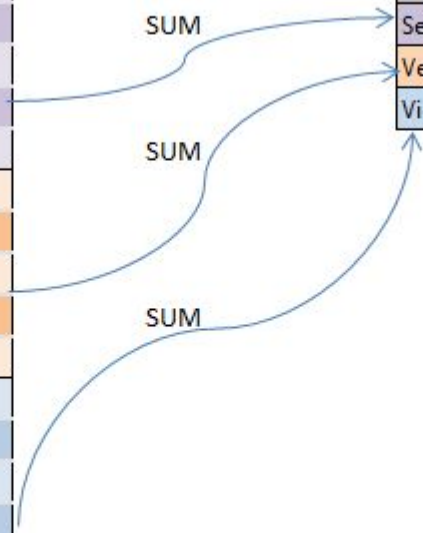
- `select()` : pick variables by their names



- `mutate()`: create new variables with functions of existing variables

- `summarise()`: collapse many values down to a single summary

sepal.length	sepal.width	variety
5.1	3.5	Setosa
4.9	3	Setosa
4.7	3.2	Setosa
4.6	3.1	Setosa
5	3.6	Setosa
7	3.2	Versicolor
6.4	3.2	Versicolor
6.9	3.1	Versicolor
5.5	2.3	Versicolor
6.5	2.8	Versicolor
6.3	3.3	Virginica
5.8	2.7	Virginica
7.1	3	Virginica
6.3	2.9	Virginica
6.5	3	Virginica
7.6	3	Virginica
4.9	2.5	Virginica



variety	sepal.length	sepal.width
Setosa	24.3	16.4
Versicolor	32.3	14.6
Virginica	44.5	20.4

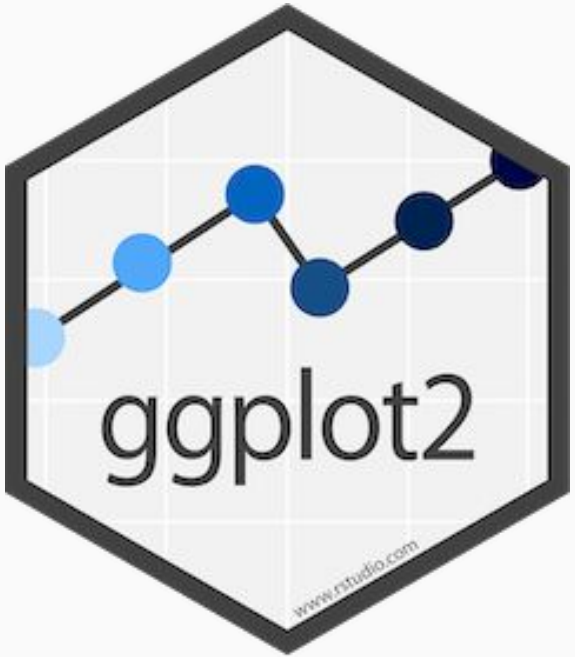
مراجعة الواجب

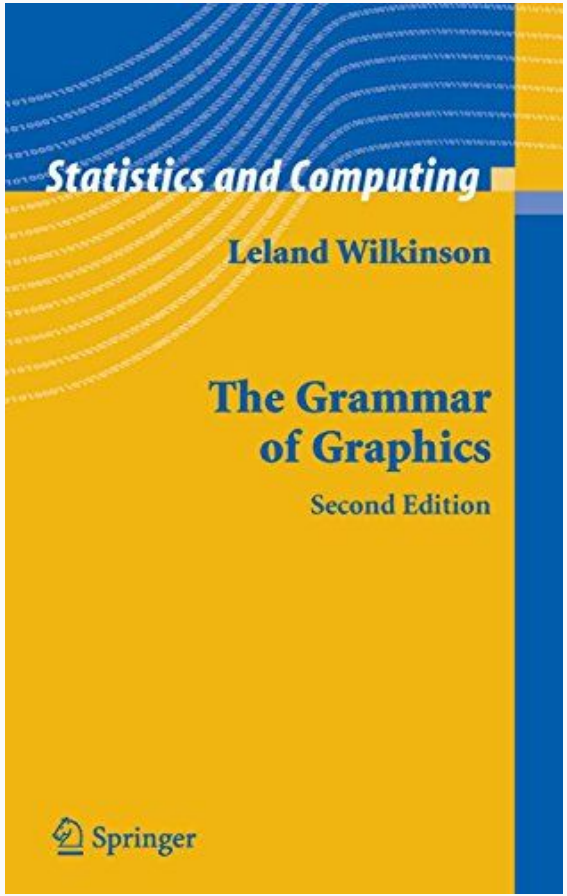
- **How many households in the survey have an average of two meals per day? Three meals per day? Are there any other numbers of meals represented?**

```
interviews %>%
  group_by(no_meals) %>%
  summarise(
     count = n()
  )
```

```
interviews %>%
  group_by(no_meals) %>%
  count()
```

Credit: Allison Horst





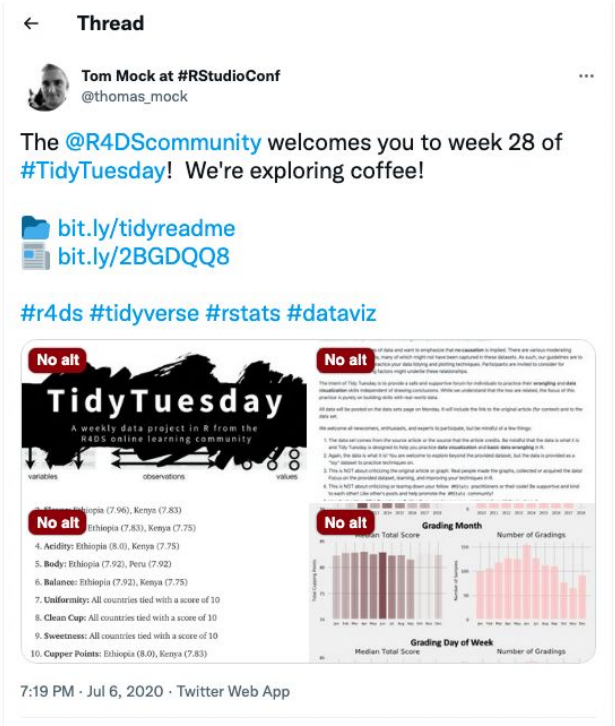
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

أمثلة من TidyTuesday

TidyTuesday هو مشروع بيانات أسبوعي يستهدف مُجتمع R الدولي وهو مدعوم من R4DS Online Learning Community و R for Data Science ، ويركز على فهم كيفية تلخيص البيانات وترتيبها باستخدام ggplot2 و tidyr و dplyr وغيرها من الأدوات في نظام لغة الآر.

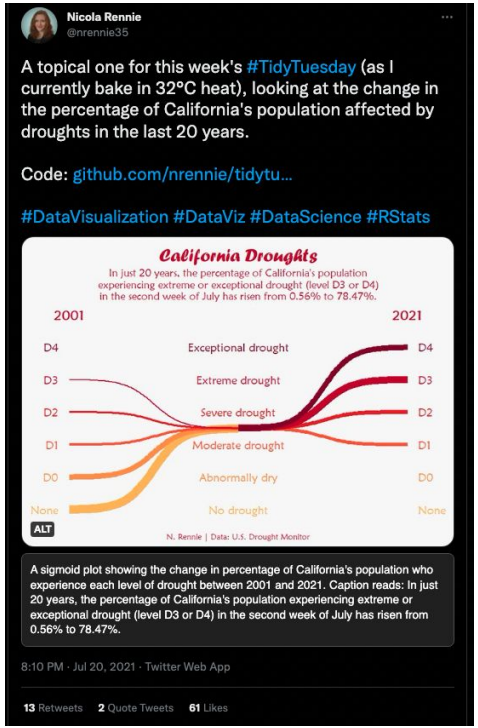
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

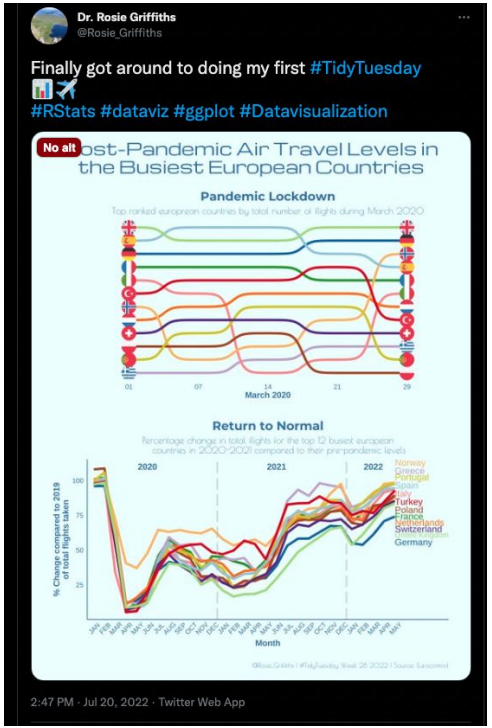
أمثلة من TidyTuesday



@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

أمثلة من TidyTuesday



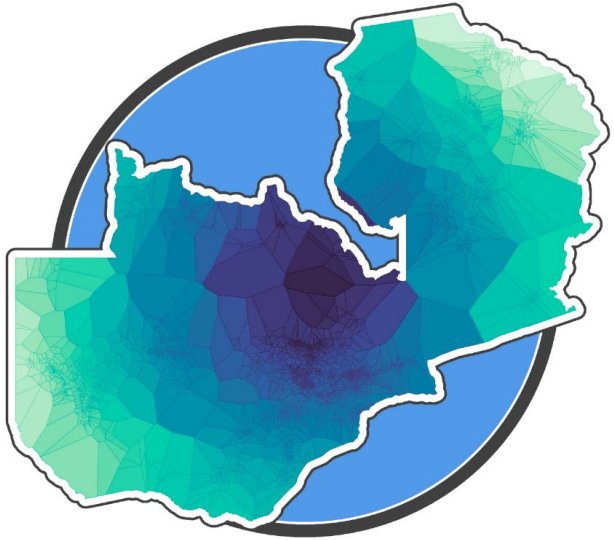


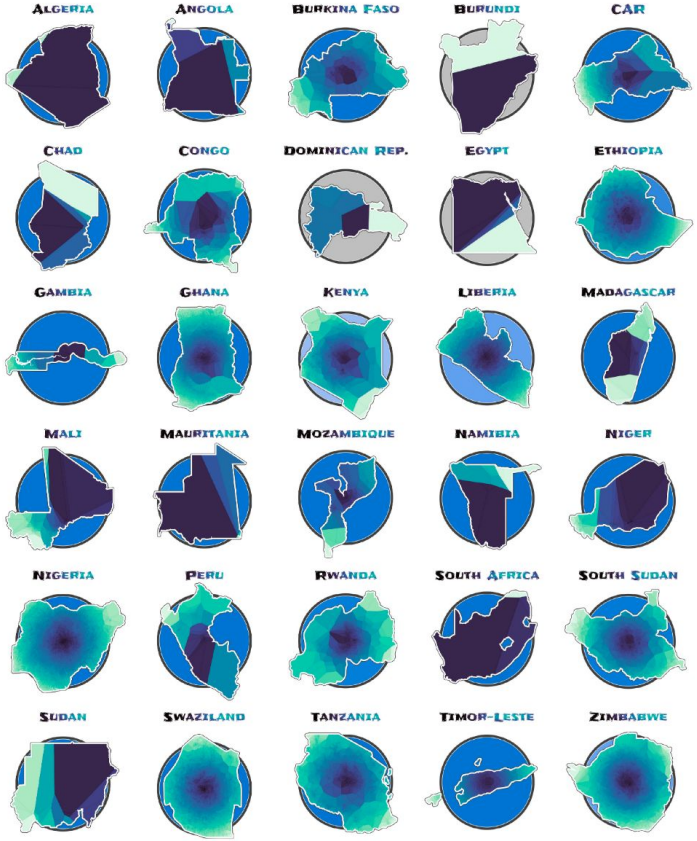
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

† I extracted all functions starting with *geom* or *stat* from my Rmd files containing the code for all my #TidyTuesday contributions (thanks Georgios for the idea and script). For the contributions from 2019 (n = 26) and 2020 (n = 40) I calculated the frequency of usage per year for each geom|stat as times used divided by the number of contributions. Note that some geom's which usually appear together (e.g. *treemapify::geomtreemap* functions) or behave very similarly (e.g. *ggforce::geommark* functions) were grouped together.

@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

ZAMBIA

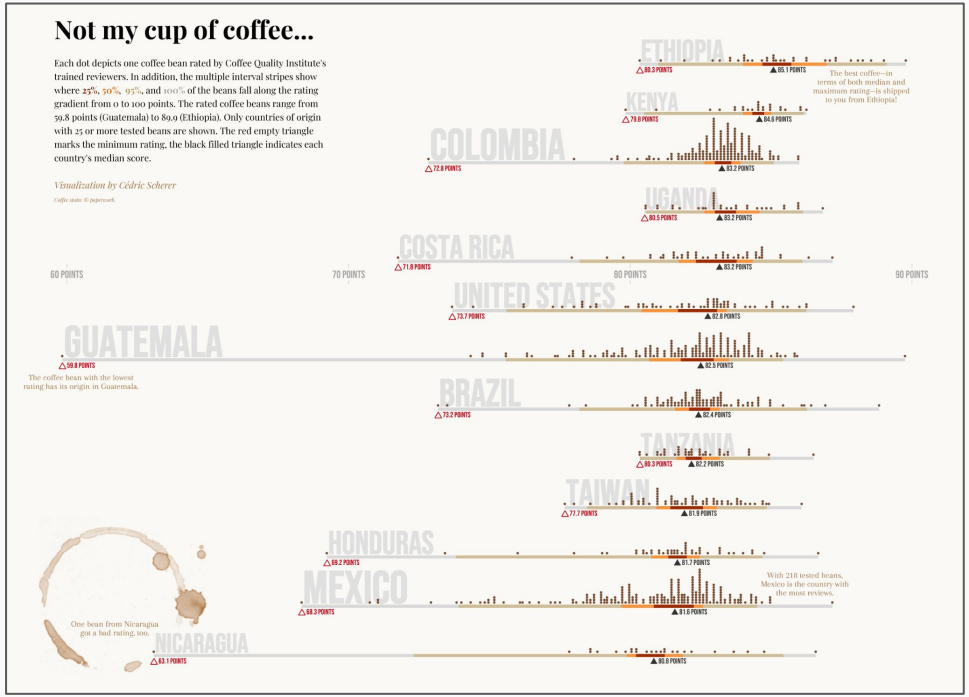


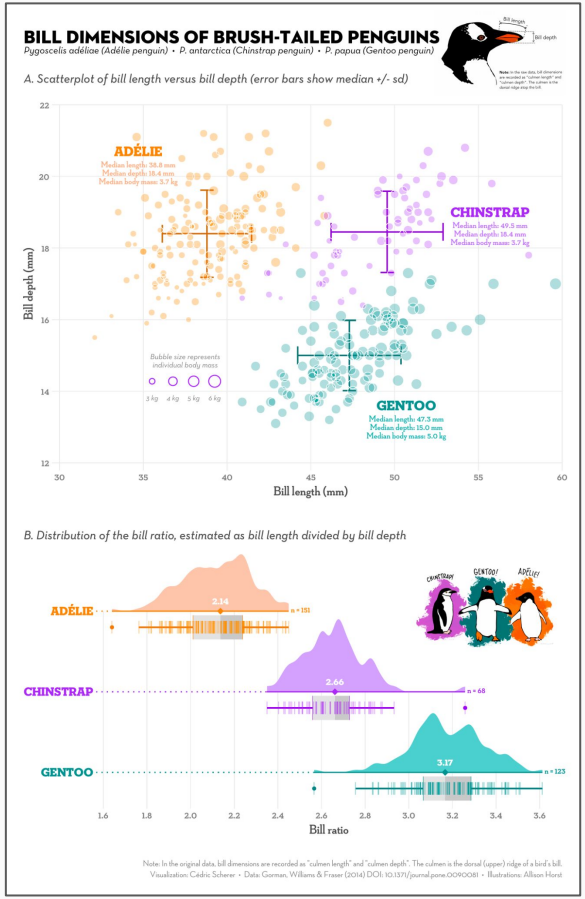


The voronoi map shows water points monitored over time by governments and development partners. The data is collected by the **Water Point Data Exchange (WPDx)** which aims to unlock the potential of water point data to improve decision-making and increase rural water access. Voronoi maps are based on the minimal distance needed to reach a landmark—*here reported water points*—by using tessellation techniques that partition a plane into regions closest to these points. The darker the color of the region, the closer it is to the country's middle point. The coloring of the circle indicates the proportion of improved versus unimproved water sources: the more intense the blue, the more improved water points; grey means no information).

Visualization: Cédric Scherer • Data: Water Point Data Exchange (WPDx), May 2021

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk





@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)
2. ربط المتغيرات بالصفات الجمالية أو الإحداثيات aes()

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)
2. ربط المتغيرات بالصفات الجمالية أو الإحداثيات aes()
3. تحديد نوع geom

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)
2. ربط المتغيرات بالصفات الجمالية أو الإحداثيات aes()
3. تحديد نوع geom

What are the building blocks in ggplot2?

1. A mapping of variables in data to
2. aes() thetic attributes of
3. geom_etric objects.

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)
2. ربط المتغيرات بالصفات الجمالية أو الإحداثيات aes()
3. تحديد نوع geom

What are the building blocks in ggplot2?

1. A mapping of variables in data to
2. aes() thetic attributes of
3. geom_etric objects.

مثال

```
ggplot(data = mydata,
aes(x =column1,
y=column2) +
geom_point())
```



@OpenSciSaudi, @batool664


www.osc-ksa.com


batool@liverpool.ac.uk

Theme

Coordinates

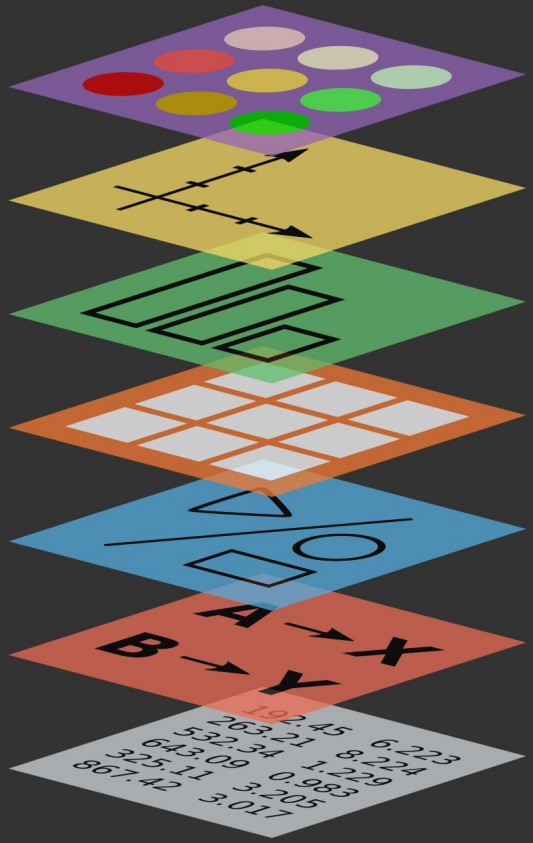
Statistics

Facets

Geometries

Aesthetics

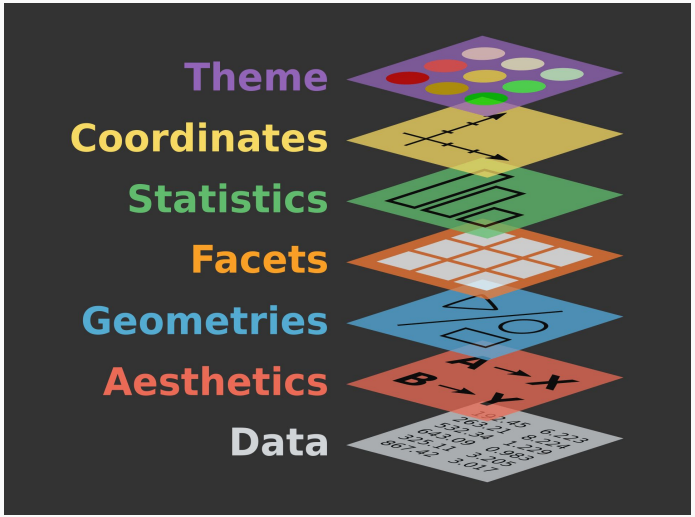
Data



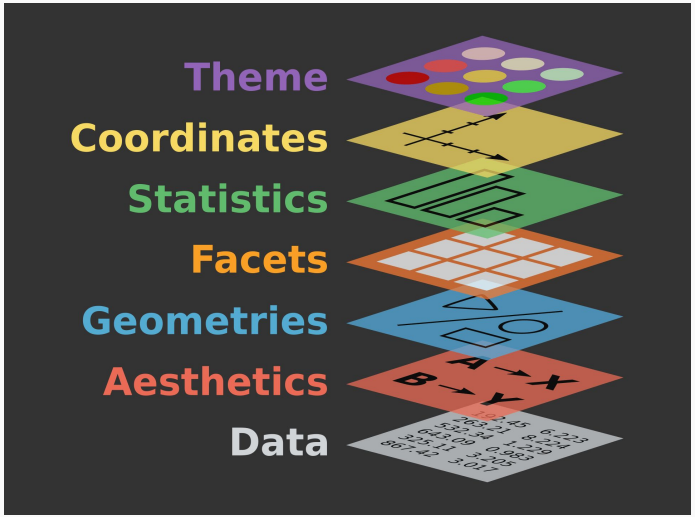
@OpenSciSaudi, @batool664

www.osc-ksa.com

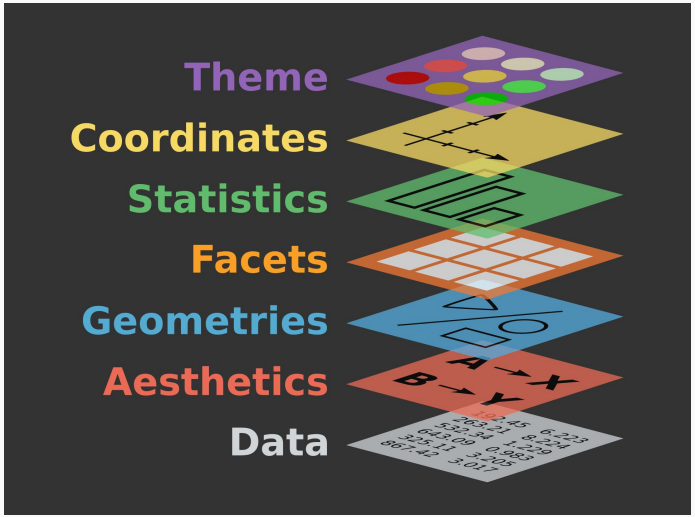
batool@liverpool.ac.uk



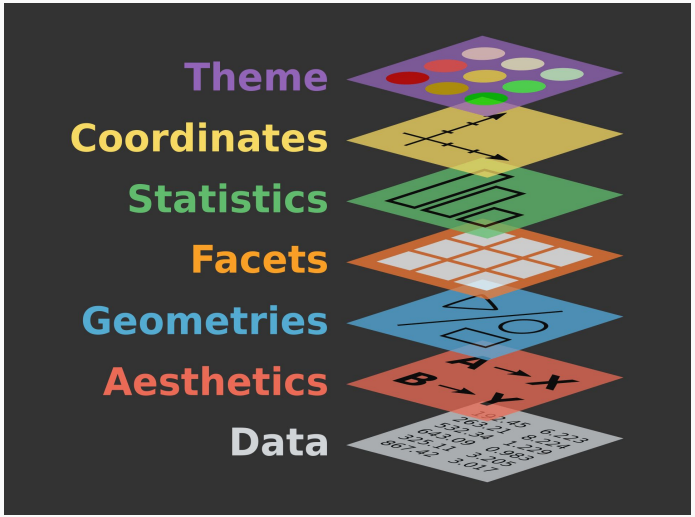

```
ggplot(data, aes(x=col1, y = col2))
```



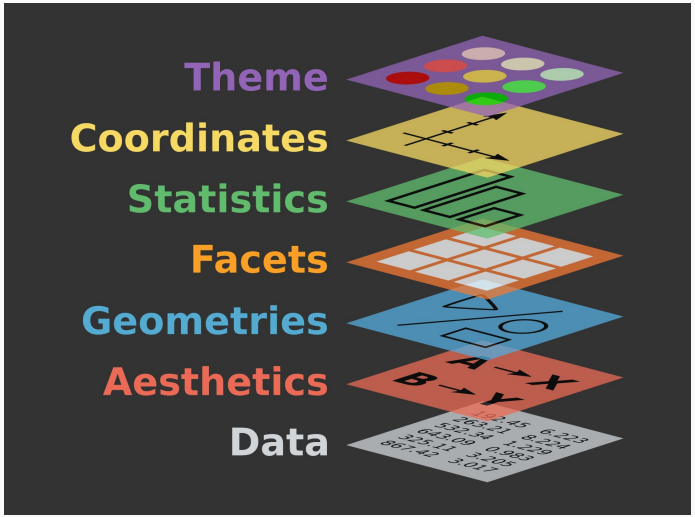

```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point()
```



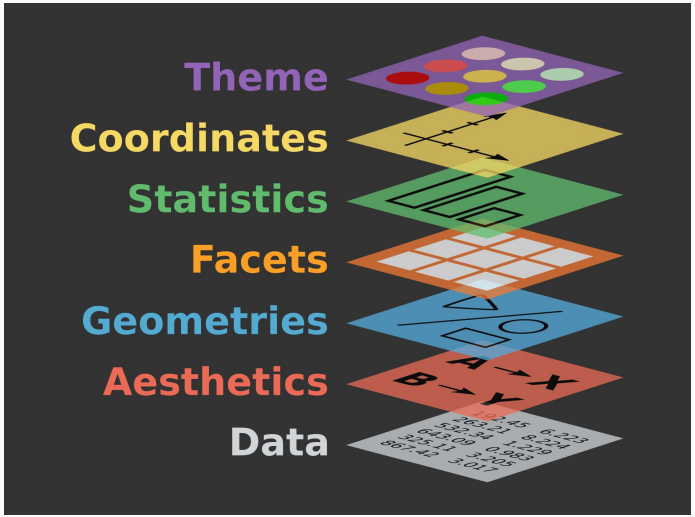

```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point() +
  geom_smooth()
```



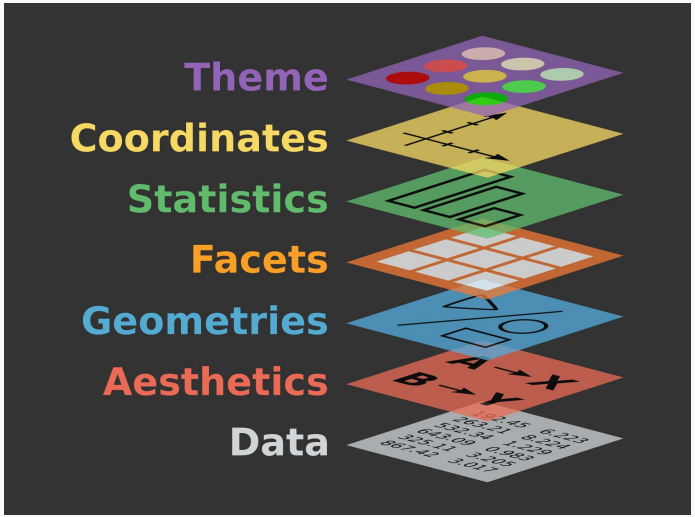

```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~ col)
```



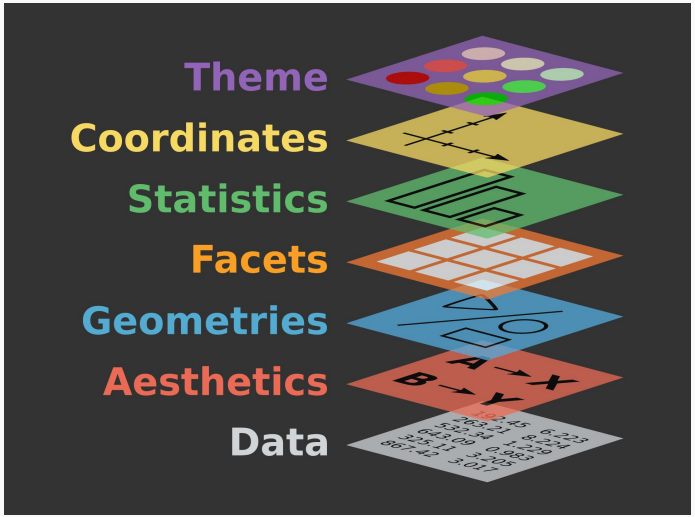

```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~ col) +
  coord_flip()
```




```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~ col) +
  coord_flip() +
  theme_bw()
```

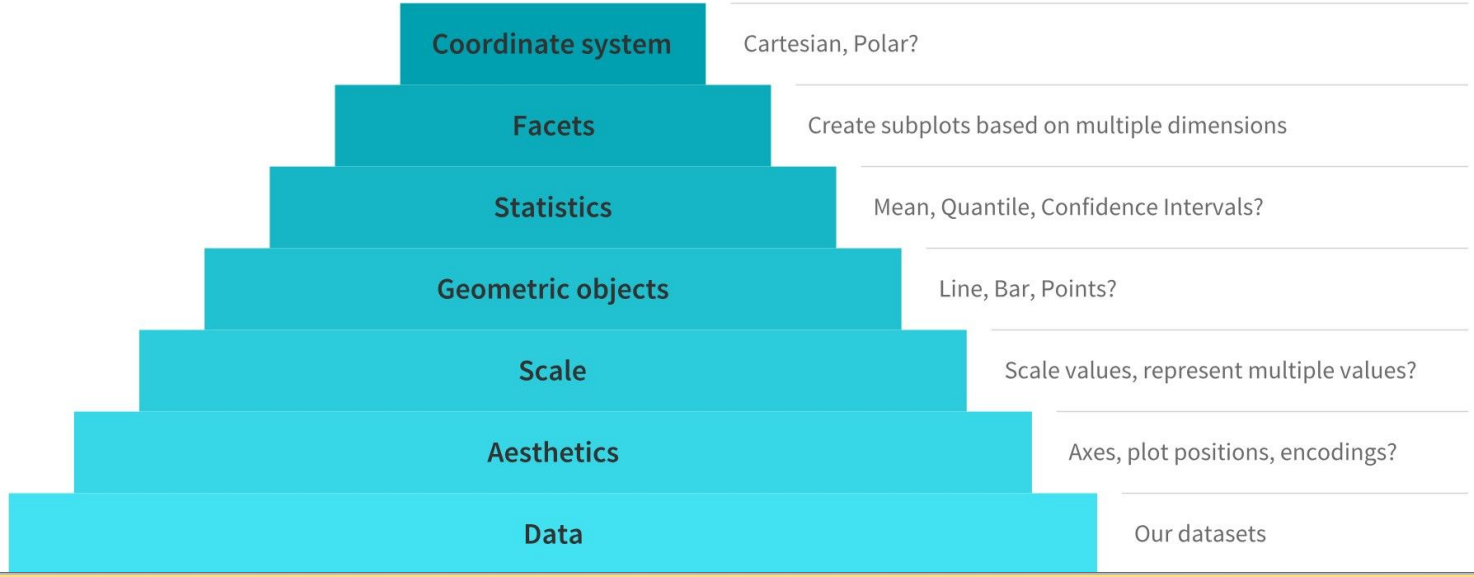



```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~ col) +
  coord_flip() +
  theme_bw() +
  theme(legend.position = "none")
```




```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point(alpha = 0.5) +
  geom_smooth() +
  facet_wrap(~ col) +
  coord_flip() +
  theme_bw() +
  theme(legend.position = "none")
```

Major Components of the Grammar of Graphics



@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

حفظ التمثيل البياني في الجهاز

```
ggsave("name_of_file.png", my_plot, width = 15, height = 10)
```

ggplot2 theme elements reference

Set `minimal` as the baseline theme:

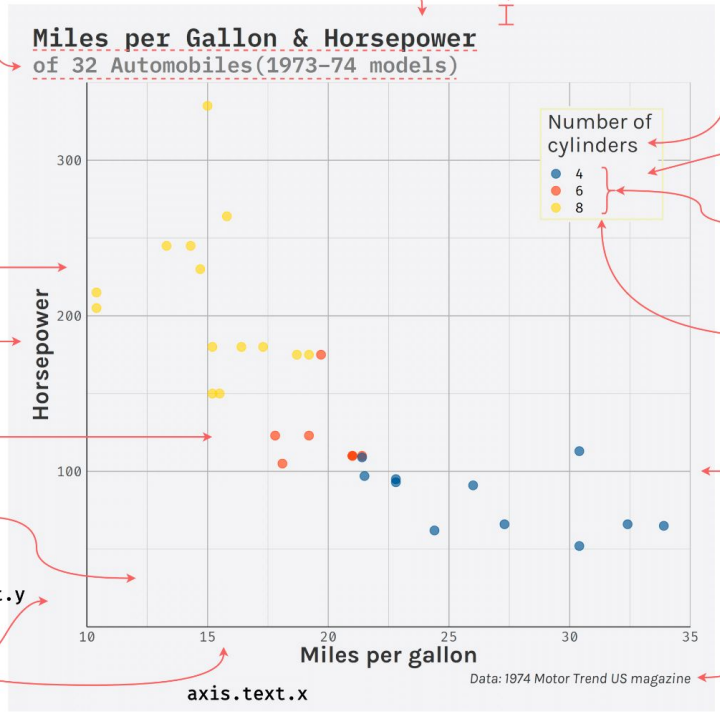
```
theme_minimal() +
theme(theme.element = element_type())
```

Use `element_blank()` to **remove an element**

Axis titles, **text**, **ticks**, and **lines** can be specified per axis using theme inheritance by putting `.x`/`.y` at the end of the theme element.

```
plot.title.position = "plot"
plot.caption.position = "plot"
```
"plot" means that they will be aligned to the entire plot (instead of the panel)

```
plot.title = element_text()
plot.subtitle = element_text()
plot.margin = margin(25, 25, 25, 25)
```




```
legend.title
= element_text()
legend.background
= element_rect()
legend.text
= element_text()
legend.position
= c(.85,.85) / "none"/
 "left"/ "right"/
 "bottom"/ "top"
plot.background
= element_rect()
plot.caption
= element_text()
axis.line.y = element_line()
axis.title.y = element_text()
panel.grid.major = element_line()
panel.grid.minor = element_line()
axis.text.y
axis.text = element_text()
axis.text.x
```

`text = element_text()` ← modifications will be applied to all text elements

Full list of elements at ggplot2.tidyverse.org/reference/theme

isabella-b

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

الواجب

- حمل حُزمة tidytuesdayR
- راجع بيانات الأسبوع ٢٨ هنا
  - https://twitter.com/thomas_mock/status/1280174453410934784
  - غرد plot جديد وسيط مبني على هذه البيانات
  - مراجعه
    - https://www.youtube.com/watch?v=-1x8Kpyndss

Feedback

- ماهو **أفضل** شيء تعلمته اليوم؟
- ماهو **أصعب** مفهوم تعرضت له في هذه الجلسة؟
- ماهو الشيء الذي تحب ان **يتحسن** في الجلسات القادمة؟