

لغة الآر من الصفر إلى الإحتراف

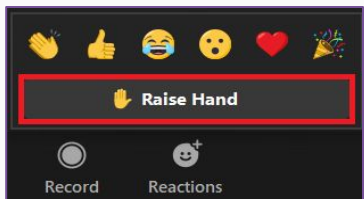
at first I was like...



Credit: Allison Horst

...but now it's like...





قواعد الدورة

● كتابة الاسم في Google Doc

○ الإجابة على Icebreaker question

● إطفاء الجوال

● يُفضل فتح الكاميرا

● لا تتردد في طرح أي سؤال خلال الجلسة

● Slack

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

المدرية: د بتول المزوق

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 وتعدديتها، 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 الاجتماعي لمن يرغب (twitter و GitHub وما إلى ذلك)

-
-
-
-
-
-
-



@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

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

كل عينة هي صف



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



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



@OpenSciSaudi, @batool664

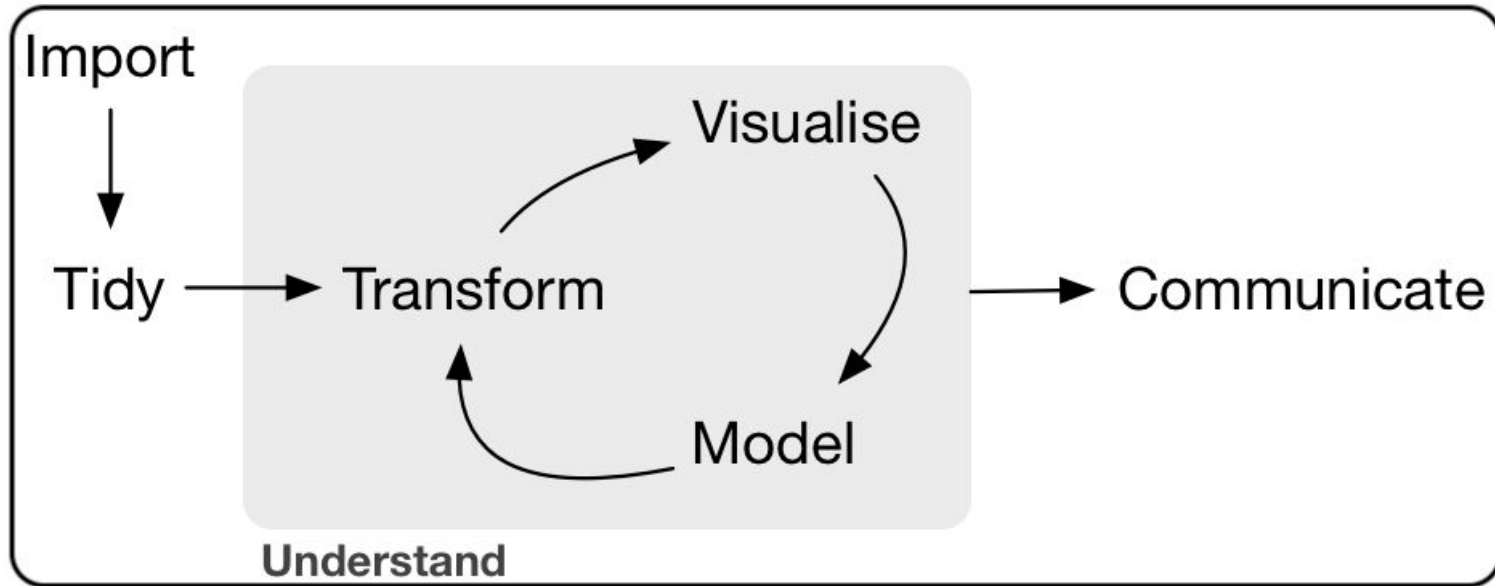


www.osc-ksa.com



batool@liverpool.ac.uk





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 تُم) = % > %

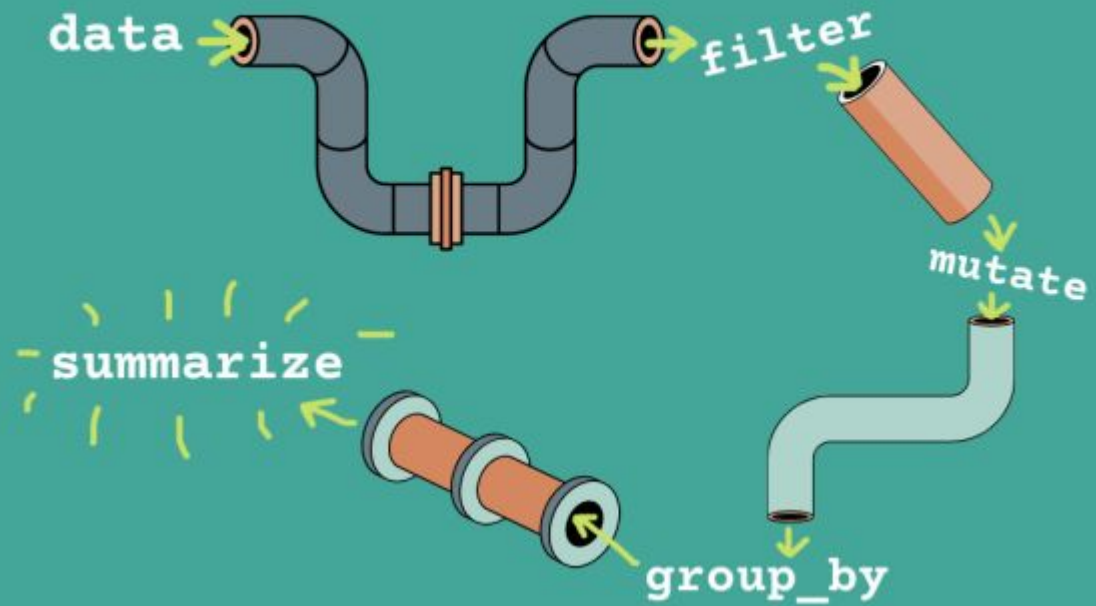
معنى هذا الرمز في لغة R

(و تُم) = %>%

dataset %>%

select(**column**)

Using pipes in



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



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

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

```
Data %>%
```

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



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk





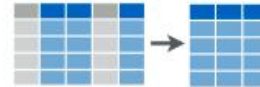
- `filter()` : pick observations by their values

Subset Observations (Rows)



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

Subset Variables (Columns)





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

Make New Variables



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

Summarise Data





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

SUM

SUM

SUM

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



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



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()
```



ggplot2: Build a data MASTERPIECE



Credit: Allison Horst

HORST '18



@OpenSciSaudi, @batool664

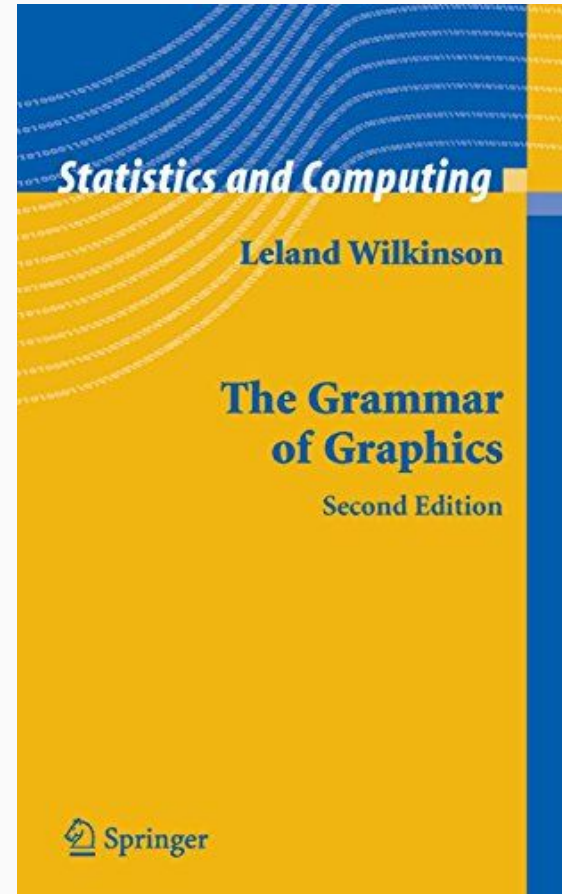
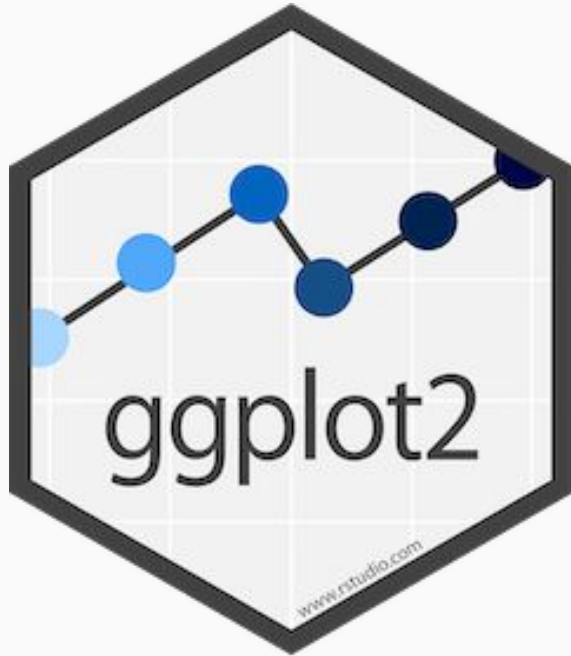


www.osc-ksa.com



batool@liverpool.ac.uk





@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 و R for Data Science و Community ، 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



← Thread



Tom Mock at #RStudioConf
@thomas_mock

The @R4DScommunity welcomes you to week 28 of #TidyTuesday! We're exploring coffee!

bit.ly/tidyreadme
bit.ly/2BGDQ08

#r4ds #tidyverse #rstats #dataviz

No alt

TidyTuesday
A weekly data project in R from the R4DS online learning community

variables → observations → values

No alt

The mission of TidyTuesday is to provide a safe and supportive forum for individuals to showcase their emerging or data visualization skills in the context of a weekly challenge. While we understand that the focus of the practice is purely on learning skills with real-world data, all data will be posted on the data sets page on Monday. It will include the link to the original article (if relevant) and the author's name.

No alt

The selection of measures, variables, and experts to participate, for the benefit of a few things:

- 1. The data sets come from the public domain or the results of open access research. We consider that the data is useful and tidy. TidyTuesday is designed to help promote data visualization and data science learning in R.
- 2. Again, the focus is on R. You are welcome to contribute to the practice in other languages, but the data is provided as a "tidy" dataset in the format of a data frame.
- 3. From the data set, you can create a plot, but please make the plot, color, and legend as clear as possible on the provided dataset, naming, and organizing your work in R.
- 4. The R4DS team will help you with any questions you have about the data set. We will not be able to help you with other data sets or help you with the R4DS community.

No alt

Grading Month

Month	Total Score	Number of Graddings
Jan	100	10
Feb	100	10
Mar	100	10
Apr	100	10
May	100	10
Jun	100	10
Jul	100	10
Aug	100	10
Sep	100	10
Oct	100	10
Nov	100	10
Dec	100	10

Grading Day of Week

Day of Week	Median Total Score	Number of Graddings
Mon	100	10
Tue	100	10
Wed	100	10
Thu	100	10
Fri	100	10
Sat	100	10
Sun	100	10

1. Ethiopia (7.96), Kenya (7.83)
2. Ethiopia (7.83), Kenya (7.75)
3. Acidity: Ethiopia (8.0), Kenya (7.75)
4. Acidity: Ethiopia (8.0), Kenya (7.75)
5. Body: Ethiopia (7.92), Peru (7.92)
6. Balance: Ethiopia (7.92), Kenya (7.75)
7. Uniformity: All countries tied with a score of 10
8. Clean Cup: All countries tied with a score of 10
9. Sweetness: All countries tied with a score of 10
10. Copper Points: Ethiopia (8.0), Kenya (7.83)

7:19 PM · Jul 6, 2020 · Twitter Web App



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



أمثلة من TidyTuesday



Nicola Rennie @nrennie35

A topical one for this week's #TidyTuesday (as I currently bake in 32°C heat), looking at the change in the percentage of California's population affected by droughts in the last 20 years.

Code: github.com/nrennie/tidytu...

#DataVisualization #DataViz #DataScience #RStats

California Droughts

In just 20 years, the percentage of California's population experiencing extreme or exceptional drought (level D3 or D4) in the second week of July has risen from 0.56% to 78.47%.

2001 2021

D4 Exceptional drought D4
D3 Extreme drought D3
D2 Severe drought D2
D1 Moderate drought D1
D0 Abnormally dry D0
None No drought None

ALT

N: Rennie | Date: U.S. Drought Monitor

A sigmoid plot showing the change in percentage of California's population who experience each level of drought between 2001 and 2021. Caption reads: In just 20 years, the percentage of California's population experiencing extreme or exceptional drought (level D3 or D4) in the second week of July has risen from 0.56% to 78.47%.

8:10 PM · Jul 20, 2021 · Twitter Web App

13 Retweets 2 Quote Tweets 61 Likes

Dr. Rosie Griffiths @Rosie_Griffiths

Finally got around to doing my first #TidyTuesday

#RStats #dataviz #ggplot #Datavisualization

No all Post-Pandemic Air Travel Levels in the Busiest European Countries

Pandemic Lockdown

Top ranked European countries by total number of flights during March 2020

01 07 14 21 28

March 2020

Return to Normal

Percentage change in total flights for the top 12 busiest European countries in 2020-2022 compared to their pre-pandemic levels

2020 2021 2022

% Change compared to 2019 of total flights taken

Month

Norway
Denmark
Portugal
Spain
Turkey
Poland
France
Switzerland
Germany

2:47 PM · Jul 20, 2022 · Twitter Web App



@OpenSciSaudi, @batool664

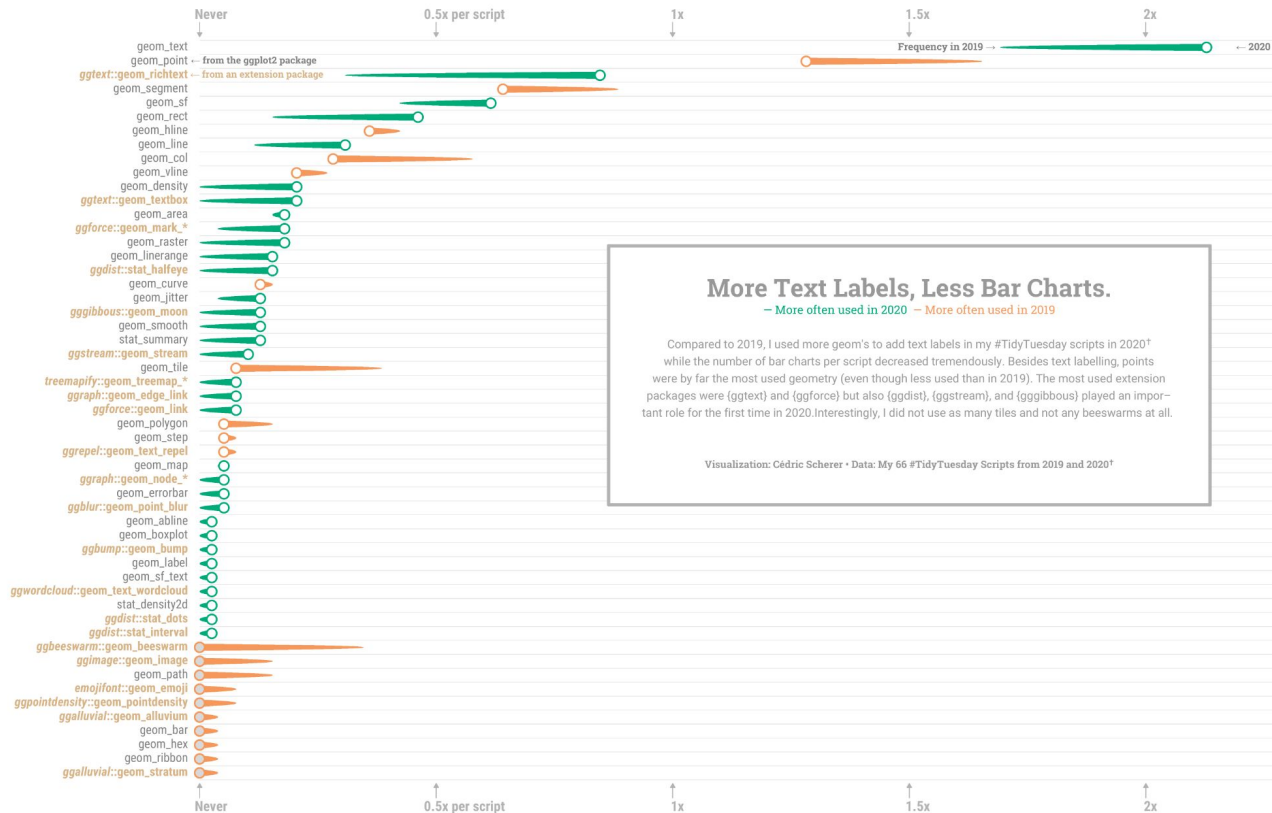


www.osc-ksa.com



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::geom_treemap` functions) or behave very similarly (e.g. `ggforce::geom_mark` functions) were grouped together.



@OpenSciSaudi, @batool664



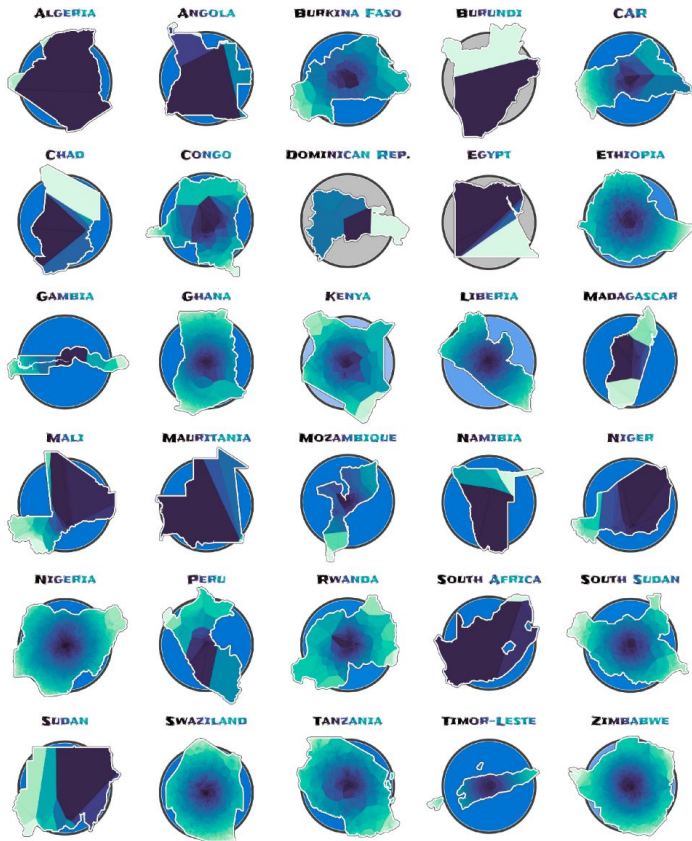
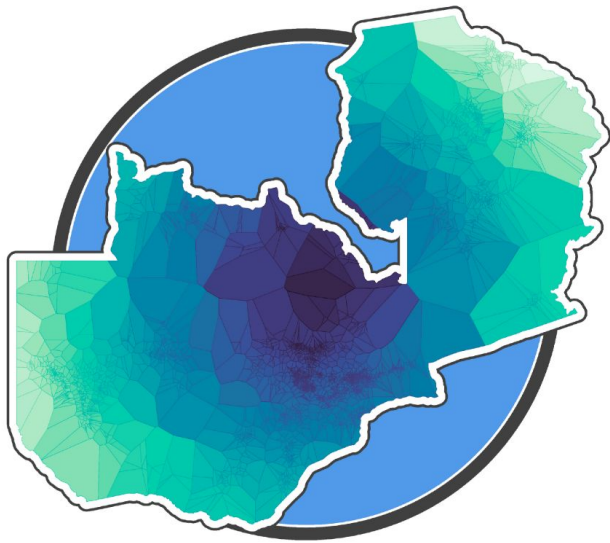
www.osc-ksa.com



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 (WPDE) 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 (WPDE), May 2021



@OpenSciSaudi, @batool664



www.osc-ksa.com



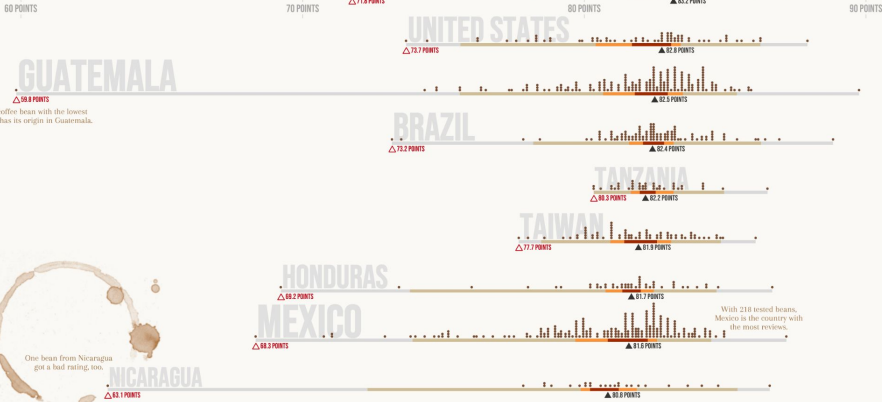
batool@liverpool.ac.uk



Not my cup of coffee...

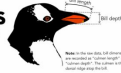
Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points. The rated coffee beans range from 59.8 points (Guatemala) to 89.9 (Ethiopia). Only countries of origin with 25 or more tested beans are shown. The red empty triangle marks the minimum rating, the black filled triangle indicates each country's median score.

Visualization by Cédric Scherer
Coffee icon: © Inepmark

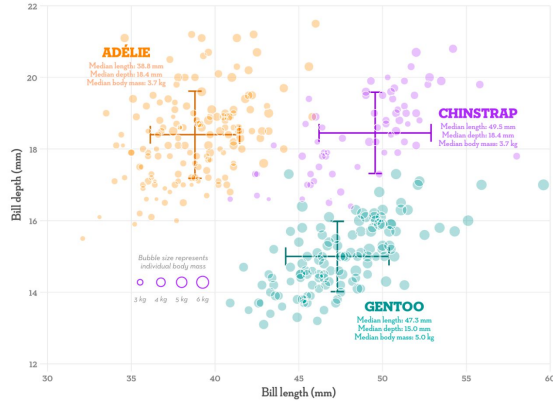


BILL DIMENSIONS OF BRUSH-TAILED PENGUINS

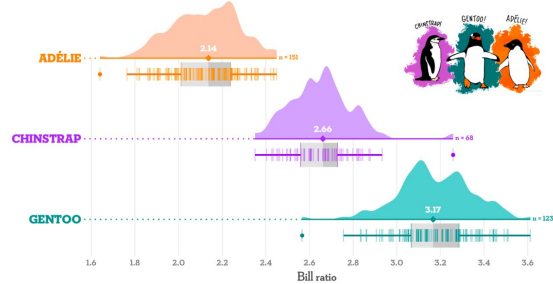
Pygoscelis adeliae (Adélie penguin) • *P. antarctica* (Chinstrap penguin) • *P. papua* (Gentoo penguin)



A. Scatterplot of bill length versus bill depth (error bars show median +/- sd)



B. Distribution of the bill ratio, estimated as bill length divided by bill depth



Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization Credits: Scherer • Data: Gorman, Williams & Fraser (2014) DOI:10.1371/journal.pone.0090081 • Illustration: Allison Hart



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk





كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)



@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()



@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() aesthetic 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() aesthetic attributes of
3. geom_etric objects.

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

مثال



@OpenSciSaudi, @batool664



www.osc-ksa.com

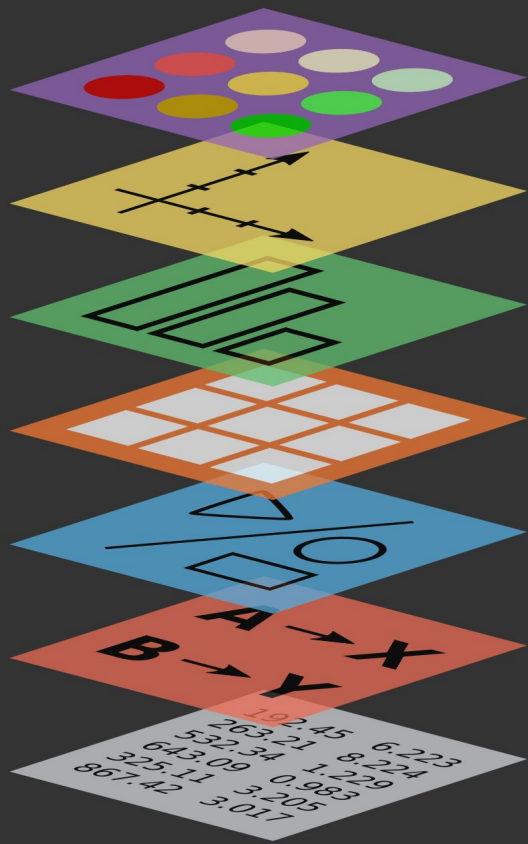


batool@liverpool.ac.uk





Theme
Coordinates
Statistics
Facets
Geometries
Aesthetics
Data



192.45	6.223
263.21	8.224
532.34	1.229
643.09	0.983
325.11	3.205
867.42	3.017



@OpenSciSaudi, @batool664

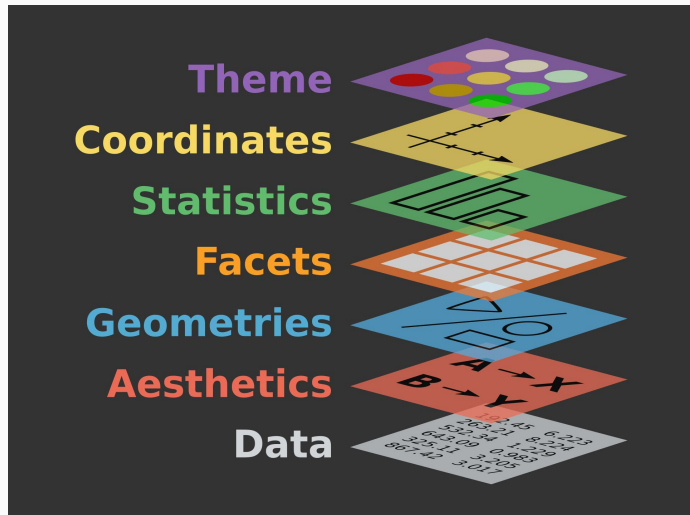


www.osc-ksa.com

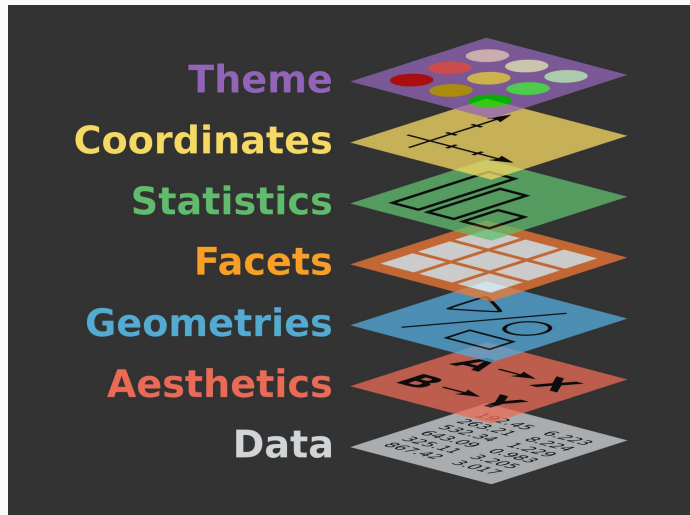


batool@liverpool.ac.uk

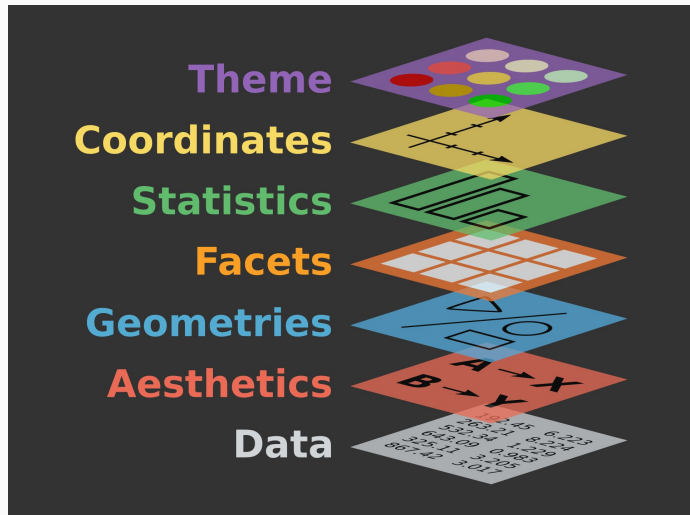




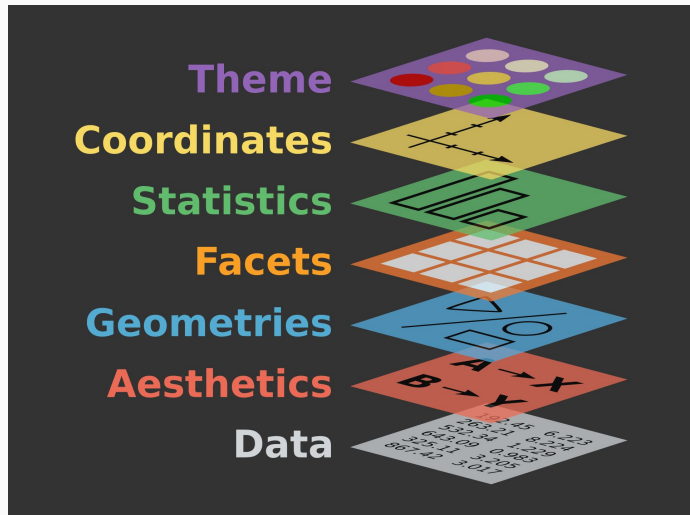
```
ggplot(data, aes(x=col1, y = col2))
```

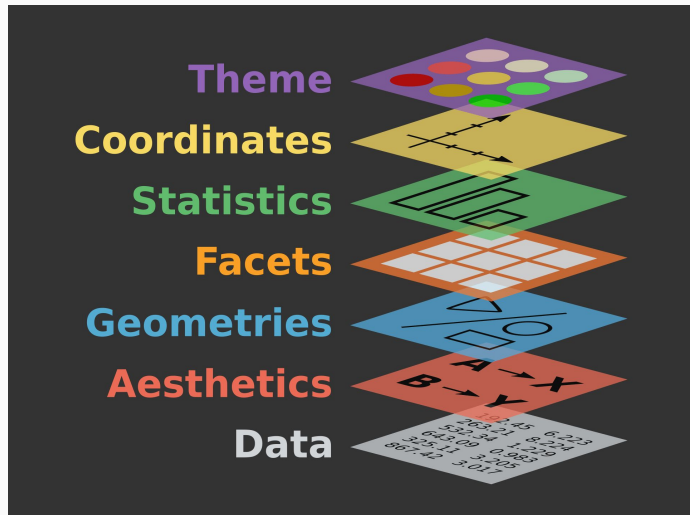
```
ggplot(data, aes(x=col1, y = col2)) +  
  geom_point()
```



```
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()
```



@OpenSciSaudi, @batool664

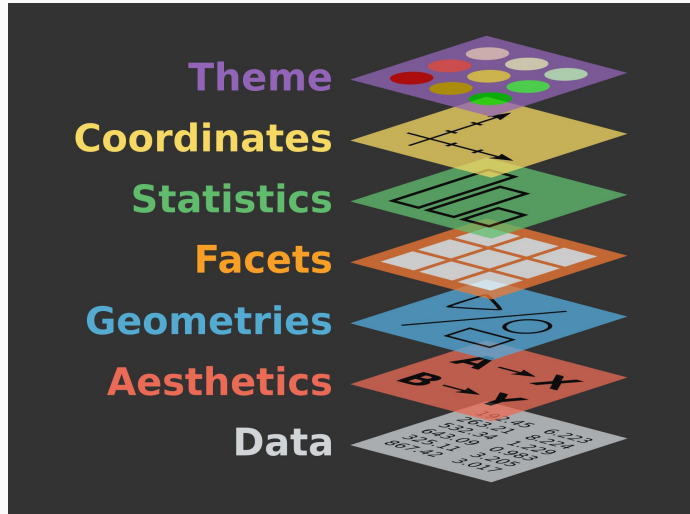


www.osc-ksa.com

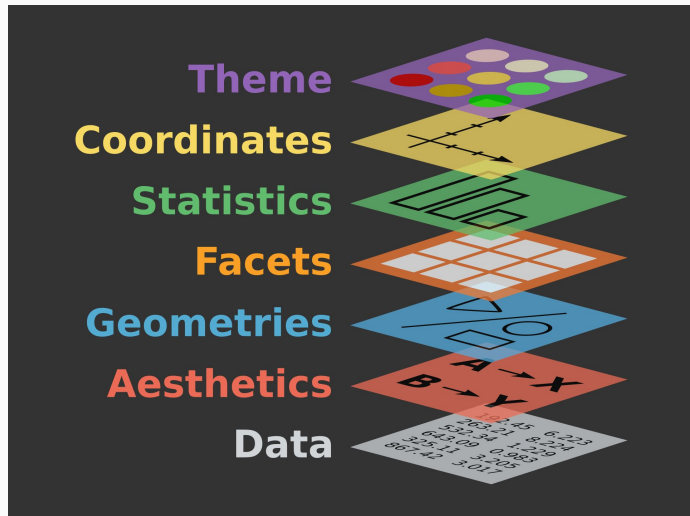


batool@liverpool.ac.uk





```
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")
```



@OpenSciSaudi, @batool664

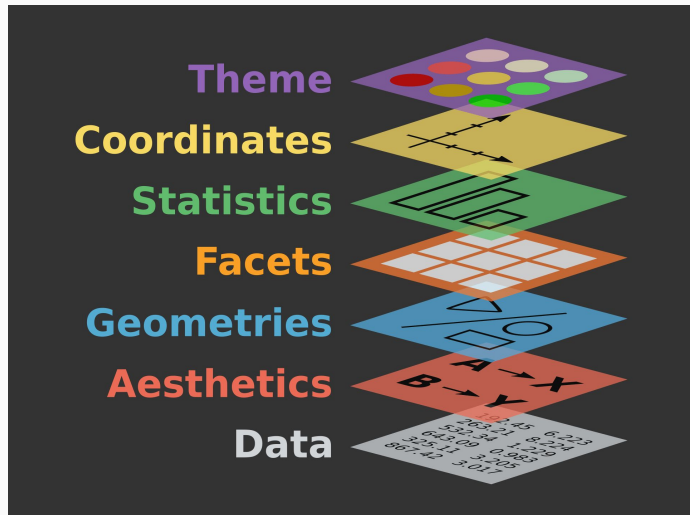


www.osc-ksa.com



batool@liverpool.ac.uk



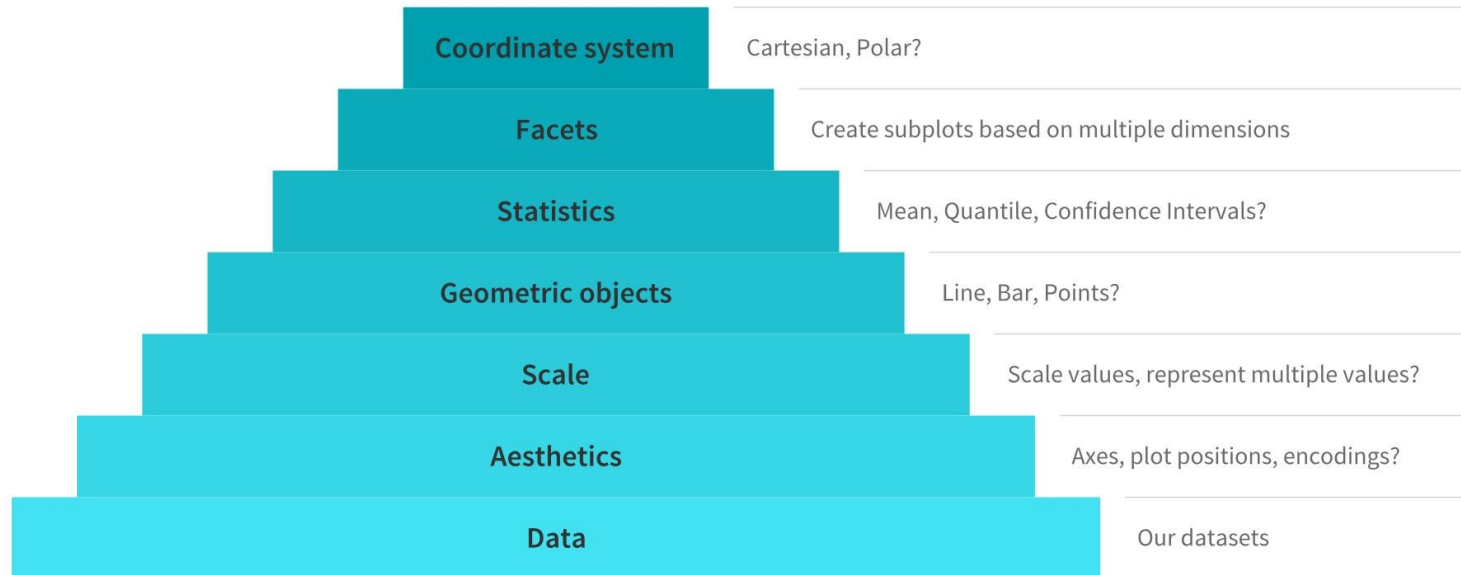


```
ggplot(data, aes(x=col1, y = col2)) +  
  geom_point(alpha = 0.5) +  
  geom_smooth() +  
  facet_wrap(~ col) +  
  coord_flip() +  
  theme_bw() +  
  theme(legend.position = "none")
```





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)
```



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk





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)
```

```
axis.line.y = element_line()
```

```
axis.title.y = element_text()
```

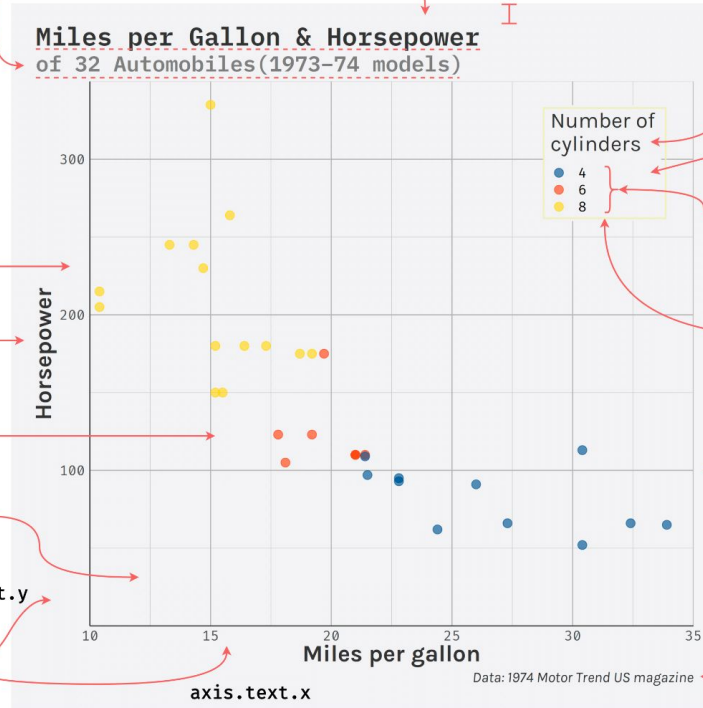
```
panel.grid.major = element_line()
```

```
panel.grid.minor = element_line()
```

```
axis.text.y
```

```
axis.text = element_text()
```

```
axis.text.x
```



```
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()
```

[isabella-b](#)

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

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



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



الواجب

● حمل حزمة tidytuesdayR

● راجع بيانات الأسبوع ٢٨ هنا

○ <https://twitter.com/tidyTuesday>

○ غرد plot جديد وسيط مبني على هذه البيانات

○ مراجعه

■ <https://www.youtube.com/watch?v=-1x8Kpyndss>



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



Feedback



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



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk

