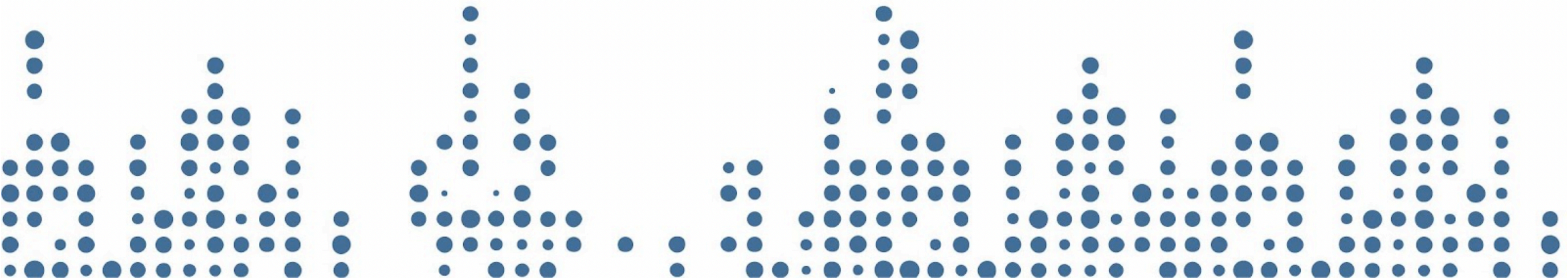




Voice-Driven Agent Assisted

Open-Source Data Science



Hello Phuse US 2026!



-  Carson Sievert @cpsievert
-  Phil Bowsher @philbowsher

GenAI in Pharma 2026: June 16th-17th

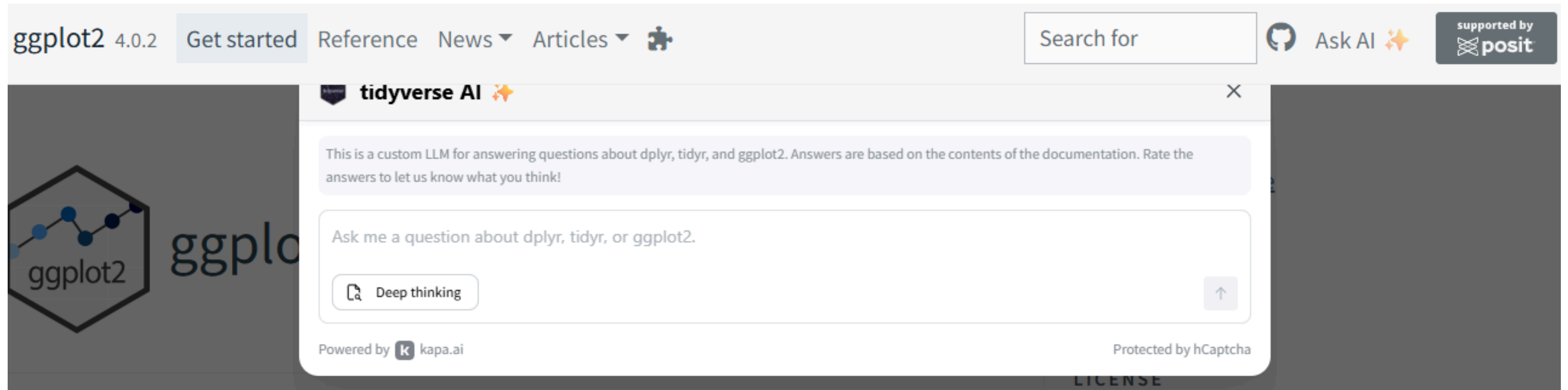




`github.com/tidyverse/ggbot`



ggplot2 is a system for declaratively creating graphics...



Posit PBC champions code-first open source data science...

- 🧑🏻‍💻 Harnessing LLMs for Data Analysis
- Can LLMs give us AGI if they are bad at arithmetic?
- Databot is not a flotation device
- 🧠 Responsible by Design: Building AI That Works With You

ggbot2 Easy Setup

-  ggbot2 requires an OpenAI API key for OpenAI's Realtime API
- Sses github.com/posit-dev/shinyrealtime under the hood
- Store the API key securely in the environment
- .Renviron
-  `pak::pak("tidyverse/ggbot2")`



```
ggbot2::ggbot(mtcars)
```



ggbot2 Shiny Interface

my-r-project-ai - Positron

ggbot2

dev.workshop.posit.team/s/00c476ae9dc1b3b2faa33/p/1441d4d9/

☆

ggbot2

<

Session cost: \$0.0000

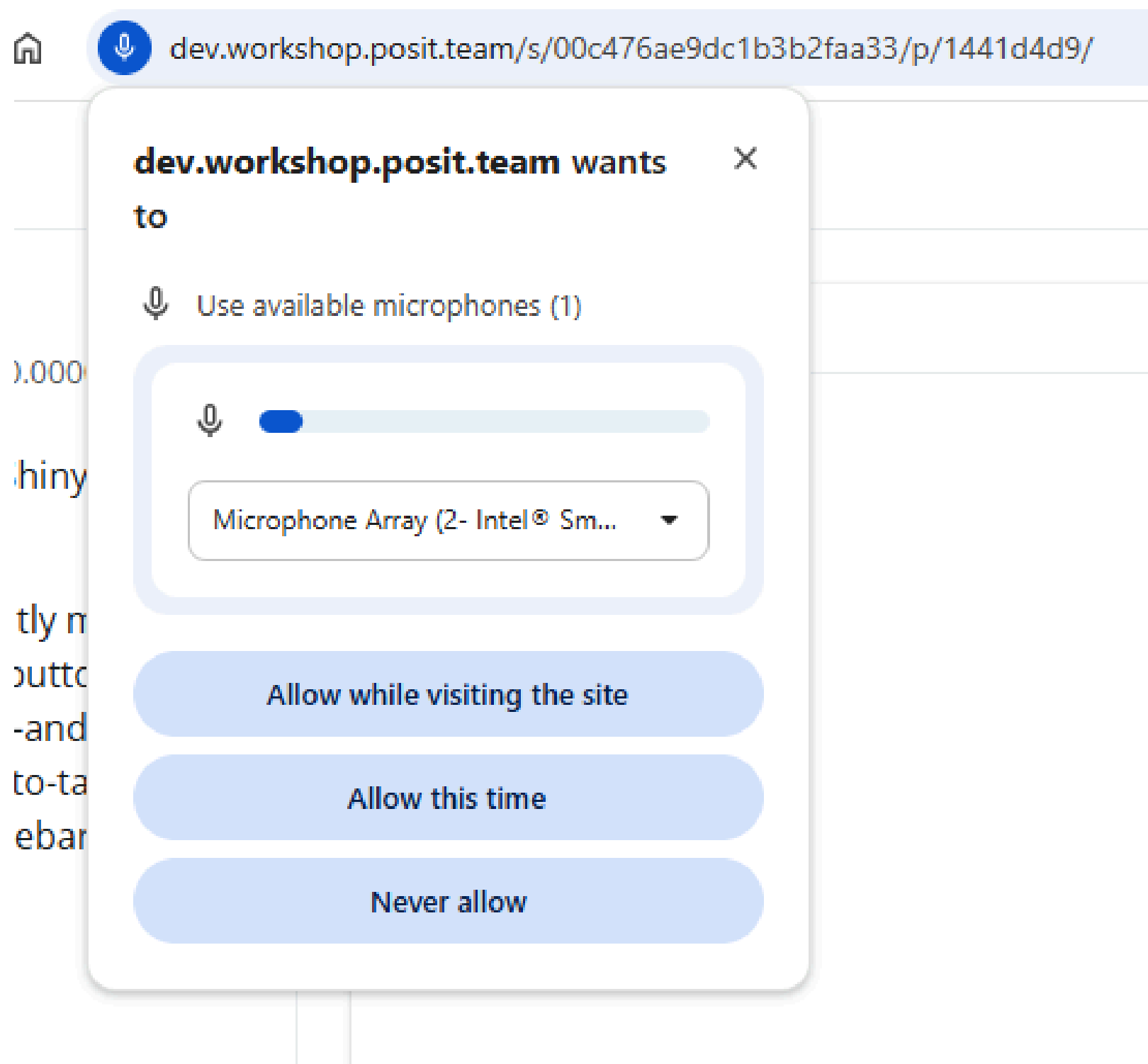
Welcome to Shiny
Realtime!

You're currently muted;
click the mic button to
unmute, click-and-hold the
mic for push-to-talk, or
hold the spacebar key for
push-to-talk.

Plot

Code

Permit a microphone



& unmute microphone





Pharma Example: ggbot2 with ADAM data

```
1 library(pharmaverseadam)
2 library(tern)
3 library(dplyr)
4
5 adsl <- adsl %>%
6   df_explicit_na()
7
8 adae <- adae %>%
9   df_explicit_na()
```



Load the 2 datasets into the Session

Search

R 4.4.0

my-r-project-ai

Data: adsl

Column Sorting

Convert to Code

Sort by Original

Filter

> A STUDYID

0%

> A USUBJID

0%

> A SUBJID

0%

> A RFSTDTC

0%

> A RFENDTC

0%

> A RFXSTDTC

0%

> A RFXENDTC

0%

> A RFICDTC

0%

> A RFPENDTC

0%

> A DTHDTC

0%

	STUDYID fct(1)	USUBJID fct(306)	SUBJID fct(306)	RF fct
1	CDISCPIL0T01	01-701-1015	1015	20
2	CDISCPIL0T01	01-701-1023	1023	20
3	CDISCPIL0T01	01-701-1028	1028	20
4	CDISCPIL0T01	01-701-1033	1033	20
5	CDISCPIL0T01	01-701-1034	1034	20
6	CDISCPIL0T01	01-701-1047	1047	20
7	CDISCPIL0T01	01-701-1057	1057	<M
8	CDISCPIL0T01	01-701-1097	1097	20
9	CDISCPIL0T01	01-701-1111	1111	20
10	CDISCPIL0T01	01-701-1115	1115	20
11	CDISCPIL0T01	01-701-1118	1118	20
12	CDISCPIL0T01	01-701-1130	1130	20
13	CDISCPIL0T01	01-701-1133	1133	20

306 rows 54 columns

SESSION

CONNECTIONS

HELP

VIEWER

CLAUDE CODE

VARIABLES

R 4.4.0

Filter

DATA

> adae

[1191 rows x 107 columns] <tbl_df>

> adsl

[306 rows x 54 columns] <tbl_df>





ggbot2::ggbot(adage)



ggbot2 Easy Setup

-  Say “Hello” to start the session with ggbot2
- Use natural language
- No need to change the tone or mode of speaking
- ggbot2 will speak back and even suggest ideas or ask for next steps
-  No need to wait for it to complete the thought

ggbot2 Spoken Text

ggbot2



Session cost: \$0.1529

Alright, we can definitely explore the severity (AESEV) by race, sex, and adverse event term (AETERM). We could create a faceted plot to visualize this. How about we do a count of adverse events by severity, and break it down by race and sex using facets, plus color by severity? Let's get that plotted!

Example Spoken Prompts I Used

-  “Using the adverse events data, create a bar chart of severity by age”
- “Make the bars colorful using blue”
- “Add the study name to the title to the bar chart”
- “The bar chart is too busy, let’s remove the age groups”
-  “Let’s go back to the last version”

Simple Bar Chart of Severity:

dev.workshop.posit.team/s/00c476ae9dc1b3b2faa33/p/1441d4d9/

☆

ggbot2

<

Session cost: \$0.0914

Plot

Count of Adverse Events by Severity

Severity	Count
MILD	750
MODERATE	380
SEVERE	50

Code

```
library(dplyr)
library(ggplot2)

# Summarize the data to get the count of each severity level
severity_counts <- adae %>%
  count(AESEV) %>%
  arrange(n)

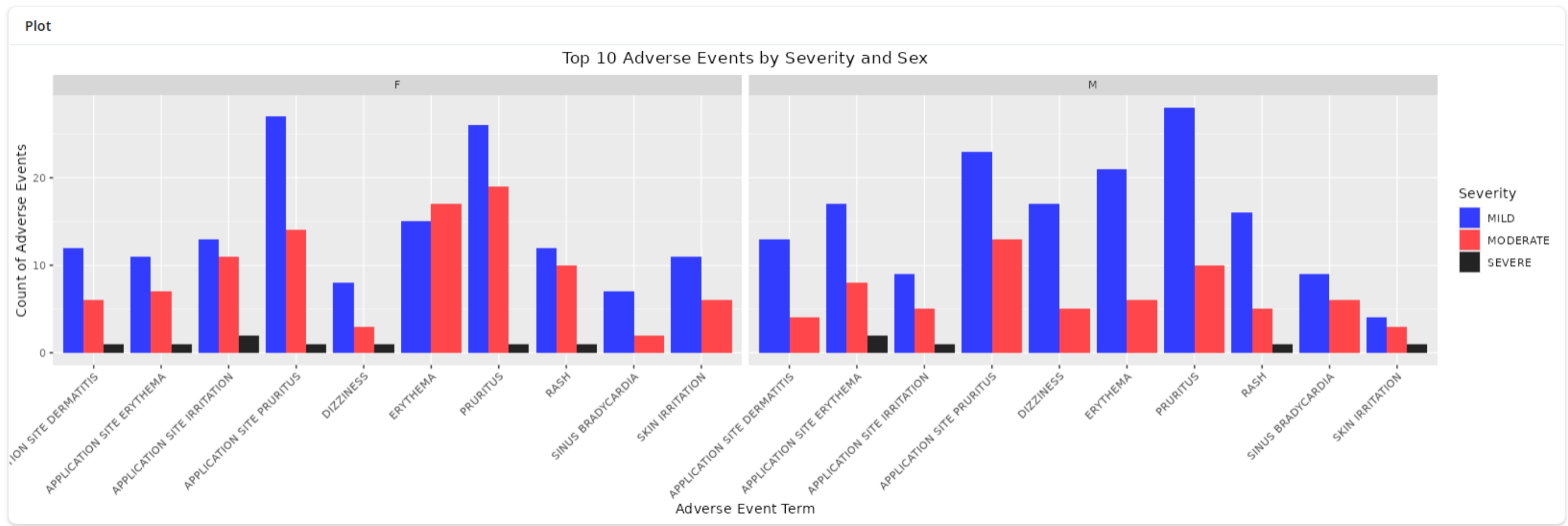
# Plot the severity counts in ascending order to see the least common first
p <- ggplot(severity_counts, aes(x = reorder(AESEV, n), y = n)) +
```

Bar chart with Adverse Events Shown:

ggbot2

Session cost: \$0.5734

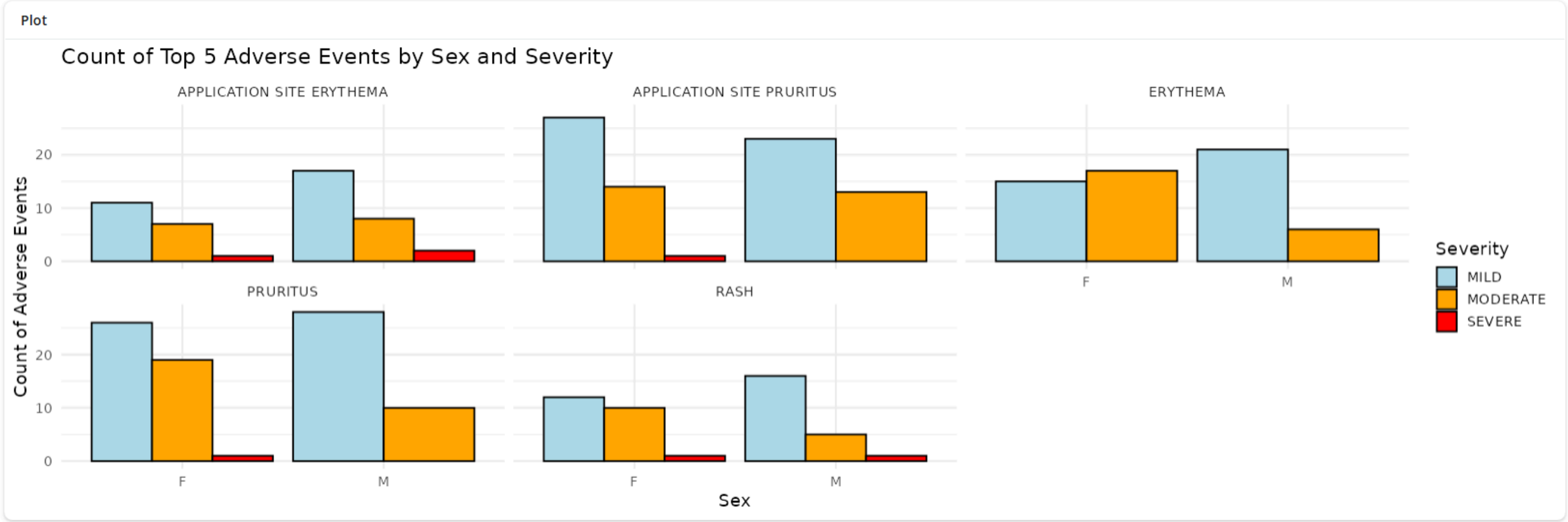
Great idea! Let's drop the race facet and instead focus on the most common adverse events. We can highlight, say, the top 10 most frequent adverse event terms and plot their severity distribution by sex. That'll give us a balanced, clearer visualization. Let's go for it!



Bar Chart Severity by Age, Sex & Top 5 Adverse Events:

ggbot2

Session cost: \$0.7605



Live Demo



0:00 / 1:44



Notes on ggbot2 Usage

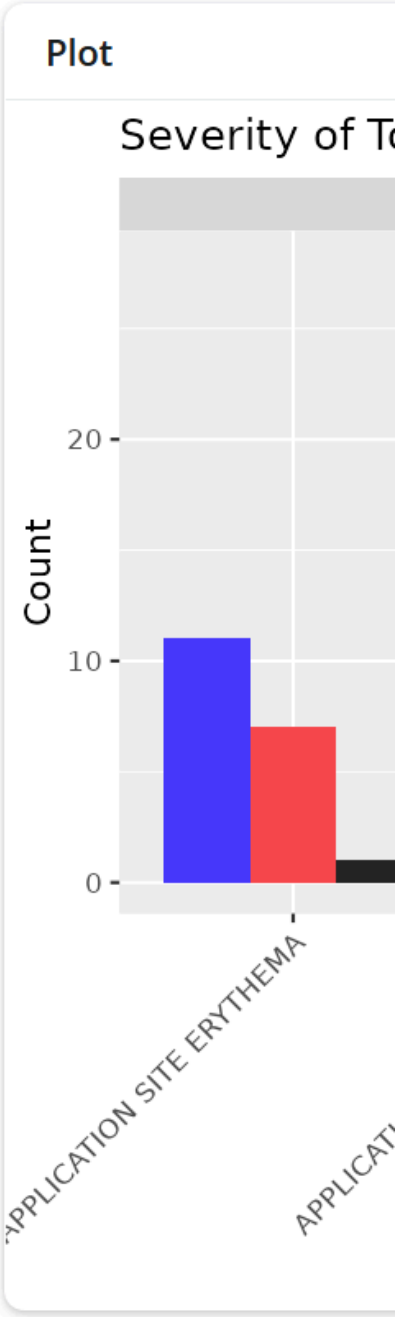
-  Packages & Python
- Cost
- R Code Reproducibility
- Missing capabilities
-  Errors - as with AI, ggbot2 can make mistakes

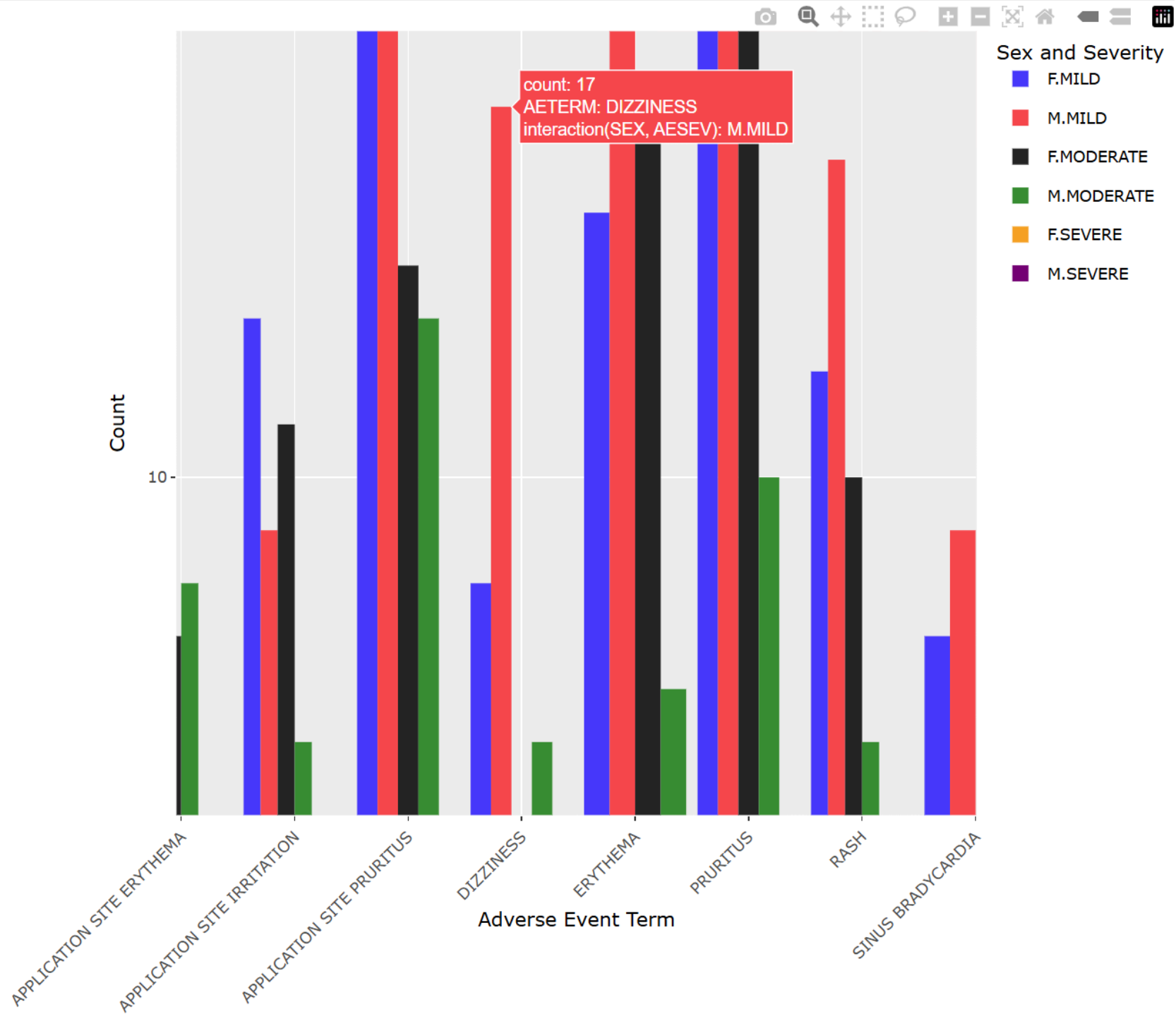
Cost

ggbot2

Session cost: \$0.2006

I'm running the plot for you right now—it'll pop up in a moment. Once we see it, we could refine it based on what stands out. For example, we could break it down further by age groups, investigate treatment arms, or add some trends over time. Let's see what the initial plot looks like first, and then we can decide the next steps together.





Error

```
library(ggplot2)
> ggbot2::ggbot(adae)
Warning: Error in count: Must group by variables found in `data`.
X Column `AGE_GROUP_10YR` is not found.
199: <Anonymous>
198: stop
197: value[[3L]]
196: tryCatchOne
195: tryCatchList
194: tryCatch
```



Future Topics



Voice-Driven Thoughts

-  Tools/apps like [Spokenly](#)
- Future state could be one where a native voice feature would be available for tools like ggbot2
- Interact with Positron, RStudio & Claude Code etc.
-  Speak in real time & revise the text to be edited & completed as needed

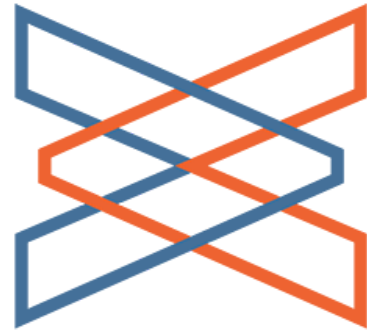
Done!

/Q: What do sharks order at McDonald's? A: A quarter\
\ flounder with cheese. /

\
\

/""-._
. '_,
: ''',
; * '.
' * () '.
\
.

Thank you!



 @philbowsher

 youtube.com/rinpharma

 rinpharma.com

