ML20: Deterministic Foundations, AI Acceleration: Lessons from Developing Hybrid Clinical Review Tools

Matthew Kumar
Srinivas Veeragoni

March 24, 2026

Background

The Challenge in Oncology Trials

// Early phase oncology trials are complex: novel compounds, unique designs, and critically ill patients.

// Time is critical—rapid data insights drive pivotal GO/No-GO decisions.

// Legacy tools are often rigid, slow, or insufficient for ad-hoc exploration.

// Traditional custom analysis requests can delay decisions by days.

// **Need**: A faster, flexible solution without compromising rigor in a regulated environment.

Our Initial Solution

A Robust, Deterministic Foundation

- We developed a self-service R Shiny application to empower clinical scientists and medical reviewers to explore baseline characteristics and efficacy.

- **Core Principle:** Provide a reliable suite of interactive tools designed to unlock maximum degrees of exploration for self-service analysis

- Built with a modular design for scalability across different studies and data models (e.g., Medical Review, ADaM)

- Key features for efficient data review and exploration:
  - Two-tier filtering (global and local scoped) system
  - Create *on-the-fly* subgroups
  - Traceable exports (PPTX, DOCX) with metadata
  - Reproducible R code generation

Data Navigator

Search Google or type a URL

Data Navigator

Study Information | Create Sub Groups | Demographic and Efficacy | Safety and Tolerability | Agentic Tools | Reports | User Guide | Bug Report | About

Bar Chart

Settings | Export | Explain

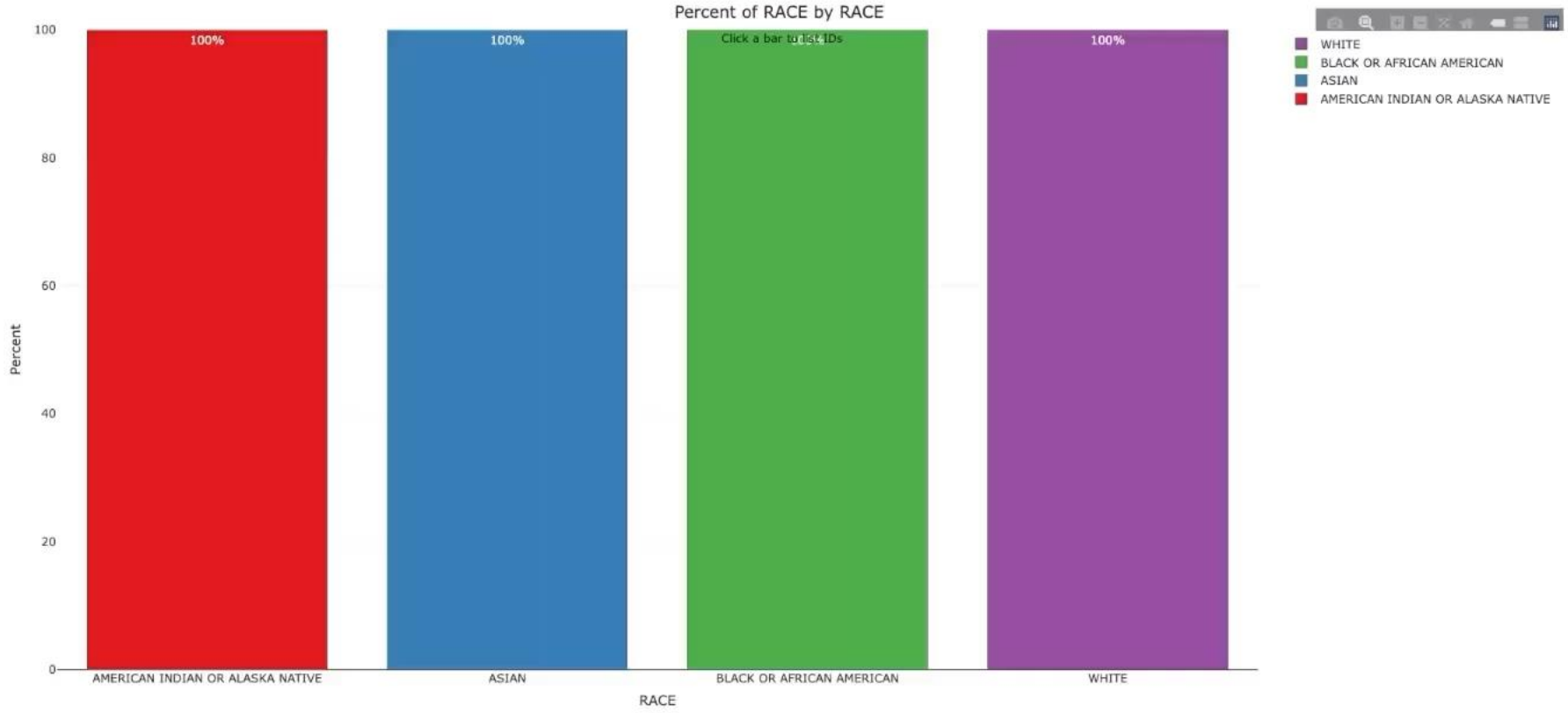
Primary Variable: Race

Stack Variable: Race

Bar Type: Stack

Display: Percent

Mode: Global



Selected Subjects

Additional Columns: Nothing selected

MESSAGE

Click a bar to see IDs.

+ Create Subgroup

Hitting a Wall

The Limits of Pre-Programmed Flexibility

// Our application was a major step forward, but we quickly saw some limitations, based on clinical scientist review workflows

// **Example**: Filtering for simple criteria (e.g., Age > 40) is easy, but dynamic ones (e.g., Filter participants above the mean of their treatment arm") were challenging.

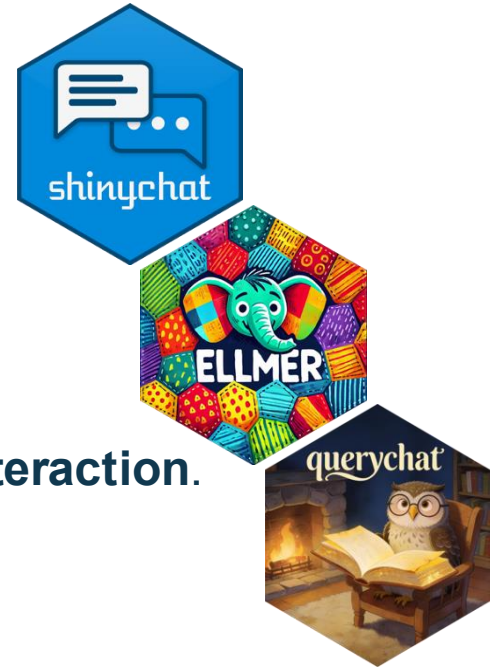
// Every unique ad-hoc request or study-specific nuance would require new, some dedicated programming.

// It was impractical to anticipate and hard-code every possible analytical question, creating a new kind of bottleneck.

// This prompted the question: can we bridge this gap through other means?

Exploring Large Language Models (LLMs)

How can we leverage them?



// Recent LLM integrations in the R ecosystem opened new possibilities.

// Implemented a chat interface using *querychat* package for natural language **data interaction**.

// Users type requests like "filter observations above the mean age of the sample"

  // System shows generated SQL query for transparency (and verification)

    // Actual data is filtered

// Data analysis capabilities for custom statistical summaries

// "Ask, review, trust but confirm" approach - quick insights with offline verification for critical decisions

Data Navigator

Study Information | Create Sub Groups | Demographic and Efficacy | Safety and Tolerability | Agentic Tools | Reports | User Guide | Bug Report | About

Bar Chart

Settings | Export | Explain

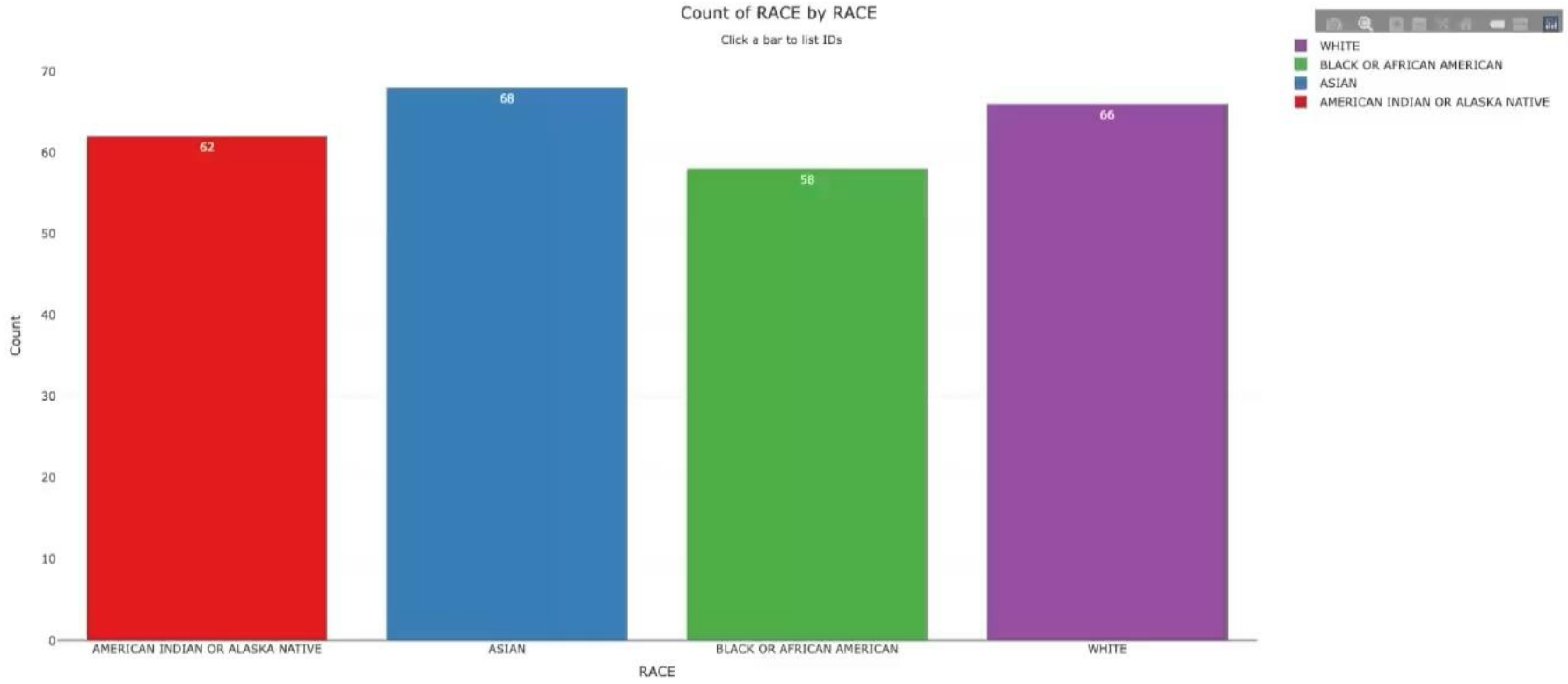
Primary Variable: Race

Stack Variable: Race

Bar Type: Stack

Display: Count

Mode: Global



Selected Subjects

Additional Columns

Nothing selected

MESSAGE

Click a bar to see IDs.

+ Create Subgroup

Expanding LLM Exploration:

Inference

// Early feedback so far: Intuitive, fast and useful. But requires user guidance for precise queries.

// LLM-assisted interpretation of visualizations/tables for rapid insights.

// Feature: One-click analysis of content like bar graphs, including user settings (e.g., filters).

// Acts as a “second pair of eyes” to spot trends or confirm observations.

// Operates in a “vacuum,” missing full study design context and nuances.

// Supports quick sanity checks while requiring human oversight for critical decisions.

Data Navigator

Study Information | Create Sub Groups | Demographic and Efficacy | Safety and Tolerability | Agentic Tools | Reports | User Guide | Bug Report | About

Box Chart

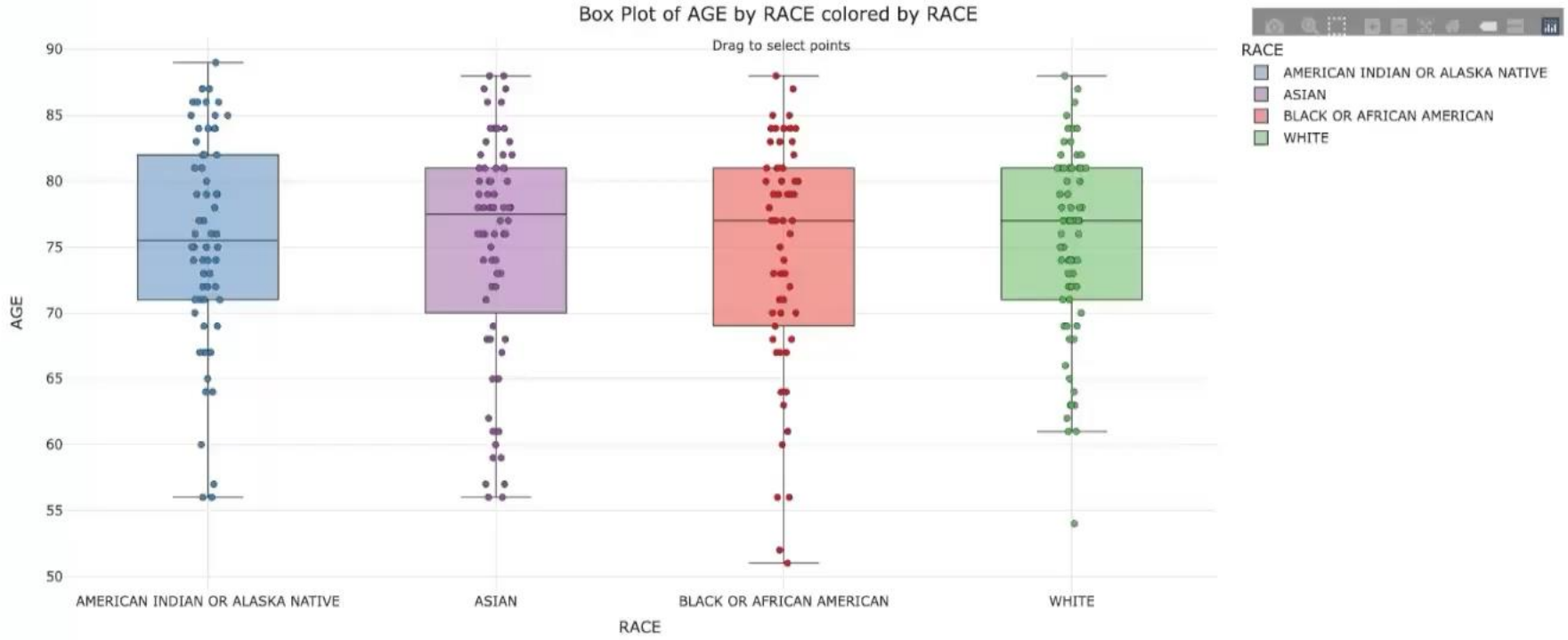
Settings | Export | Explain

Response Variable: Age

Primary Variable: Race

Secondary Variable: Race

Mode: Global



Selected Points

Additional Columns

Nothing selected

MESSAGE

No points selected. Drag on the plot to select.

+ Create Subgroup

Expanding LLM Exploration:

Novel Content

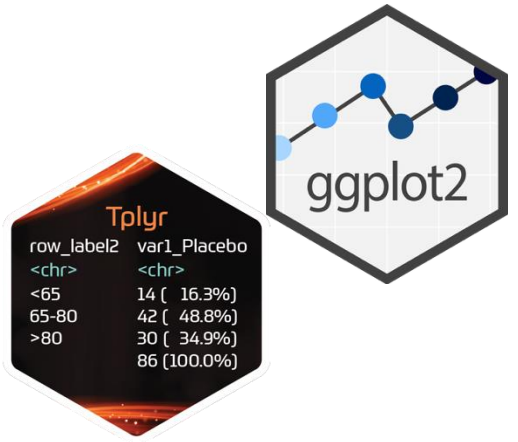
// Up to now: content is generated in deterministic, pre-programmed modules where options and filters enable flexibility (e.g. bar plot, summary table)

// Our next LLM exploration focused on generating altogether *new* plot types (ggplot2) and table summaries (Tplyr) not present anywhere in the application

// In development, and builds on the querychat approach of blending LLM tool calling and shiny reactivity.

// Positive end user reception, but with some challenges (more later).

// Examples…



Data Navigator | Study Information | Create Sub Groups | Demographic and Efficacy | Safety and Tolerability | Agentic Tools | Reports | User Guide | Bug Report | About

Data Navigator

Empowering data-driven insights for early clinical research.

0.2.9000

Study Information

Study Number	99999
Bay Number	1234567
Project Description	A Mock Compound
Study Description	A Mock Study
Phase	1
Safety Analysis Set	Y: 254 (100%) / N: 0 (0%) / Missing: 0 (0%)
Efficacy Analysis Set	Y: 234 (92.1%) / N: 20 (7.9%) / Missing: 0 (0%)

LLM Explorations:

Some Observations so far

// Complexity adds effort: Multi-step logic and nuanced operations need careful system design and testing.

// Data structure mismatches: Joining single-row (demographics) with multi-row (labs) tables failed

// Specialized clinical outputs: Novel requests (e.g., eDish plot, shift table) are difficult—LLMs may lack knowledge of them, and execution requires complex operations.

// User-jargon gaps: Terms like "faceting" or "spanning headers" aren't intuitive for clinical reviewers.

// Context awareness: Interpretations miss study design, protocol, and clinical intent.

// **Path forward**: Refining system prompts, providing additional instruction sets, and LLM evaluations can help study this in a more principled way

LLM Explorations:

Take aways

- Deterministic approaches are best for standardized, high-stakes analyses (e.g., efficacy reporting).
- LLMs excel in rapid, exploratory tasks but require verification and context. They must be utilized in the right context.
- User awareness and education on strengths and limitations of LLM technology is crucial to its effective application.
- Hybrid approach combines innovation with precision—neither alone suffices for what we aim to achieve.
  - Expose deterministic, pre-written functions as tools to LLM as a bridge

Future Looking

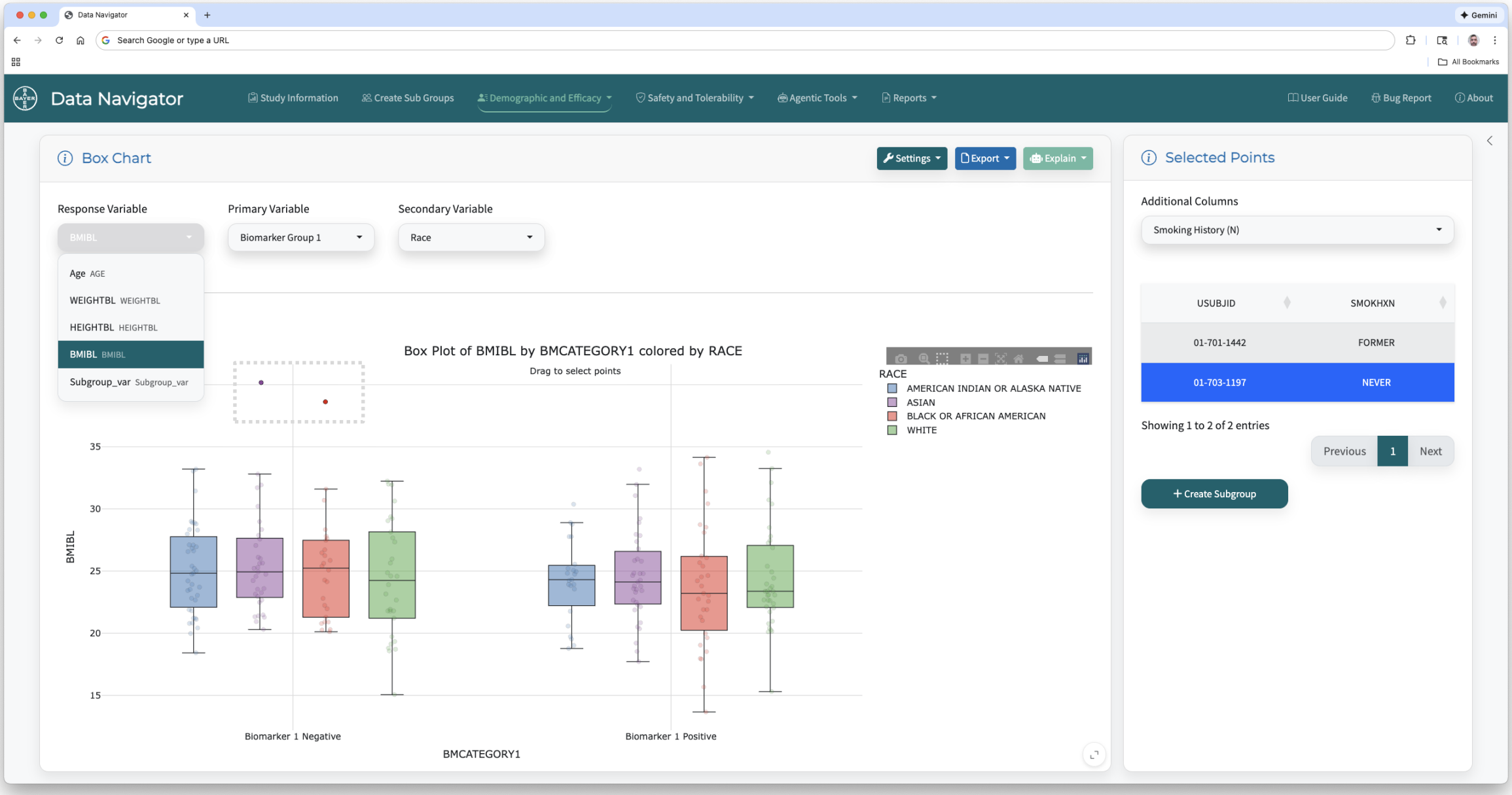
Challenges

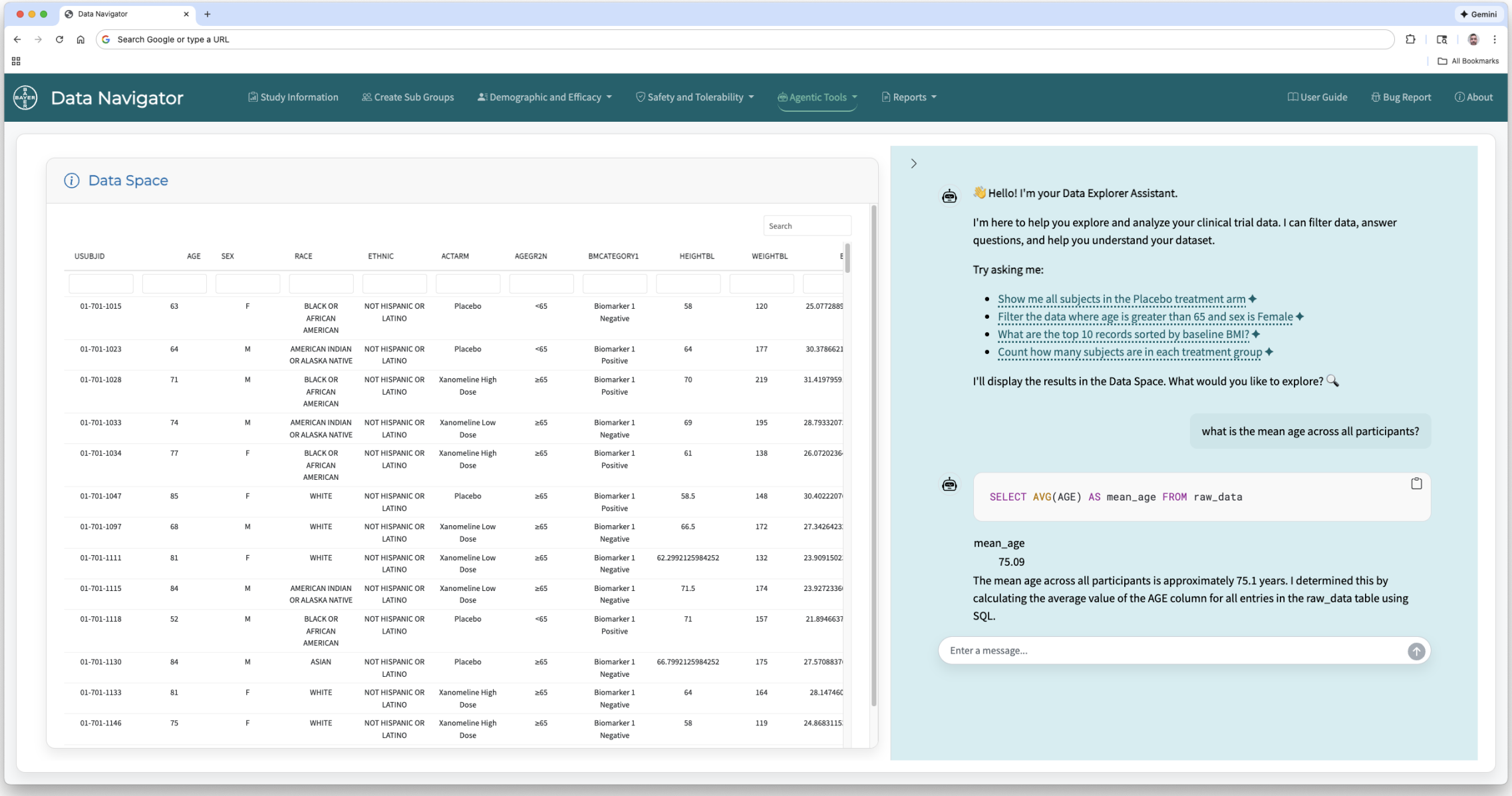
- Position our hybrid Shiny + LLM-enabled application as the cornerstone of an end-to-end disease data strategy within oncology.
- Unite all trial functions—data management, monitoring, insights, stats, and clinical leads—on one collaborative platform.
- Front-load key components from protocol and EDC building to submission reports for seamless integration.
- Enhance LLMs with context (e.g., protocol, SAP) via RAG, while maintaining HIL oversight for expert judgment.
- Transform trial conduct: Accelerate decisions, ensure rigor, and drive effective data-driven outcomes.

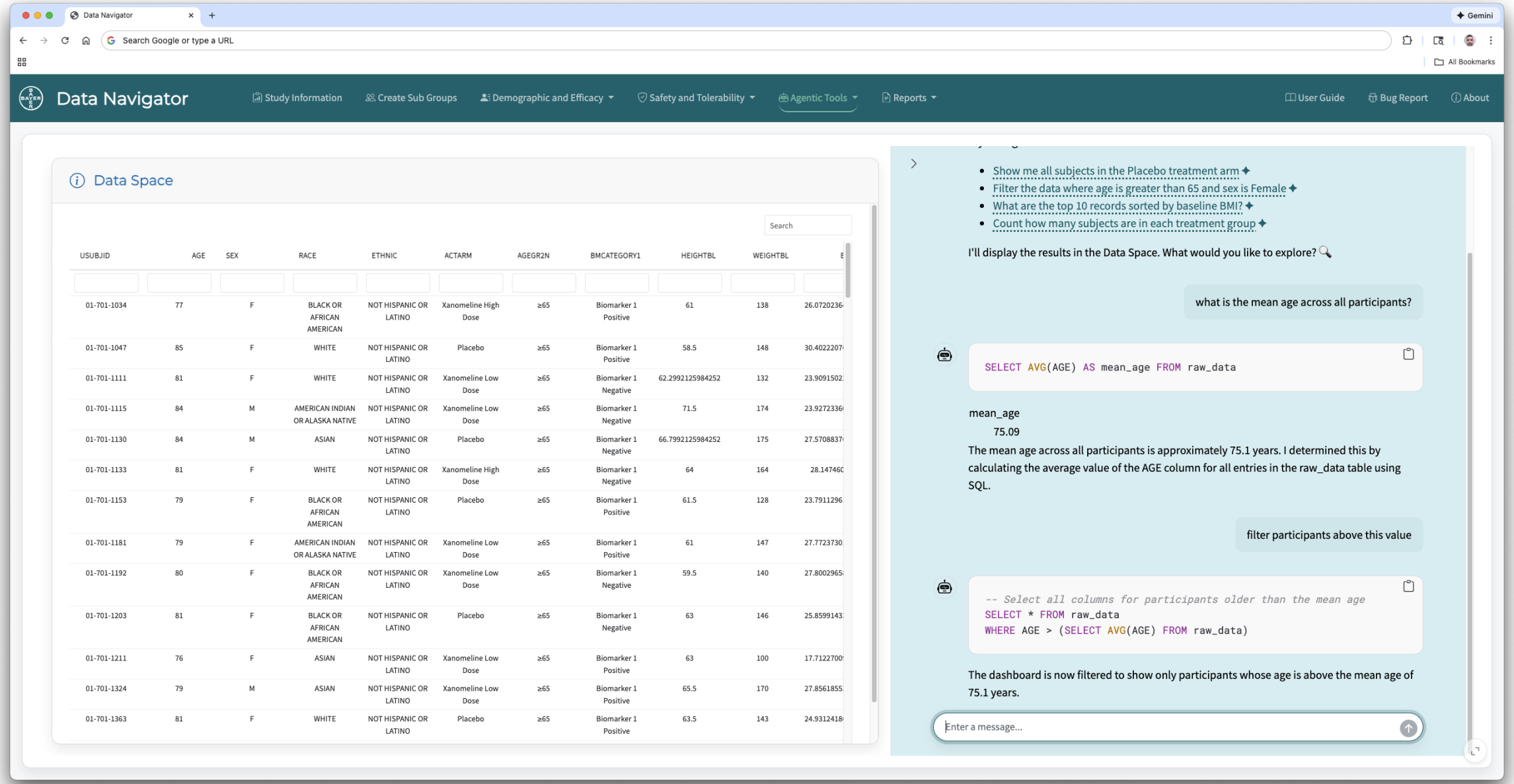
Health for all, Hunger for none

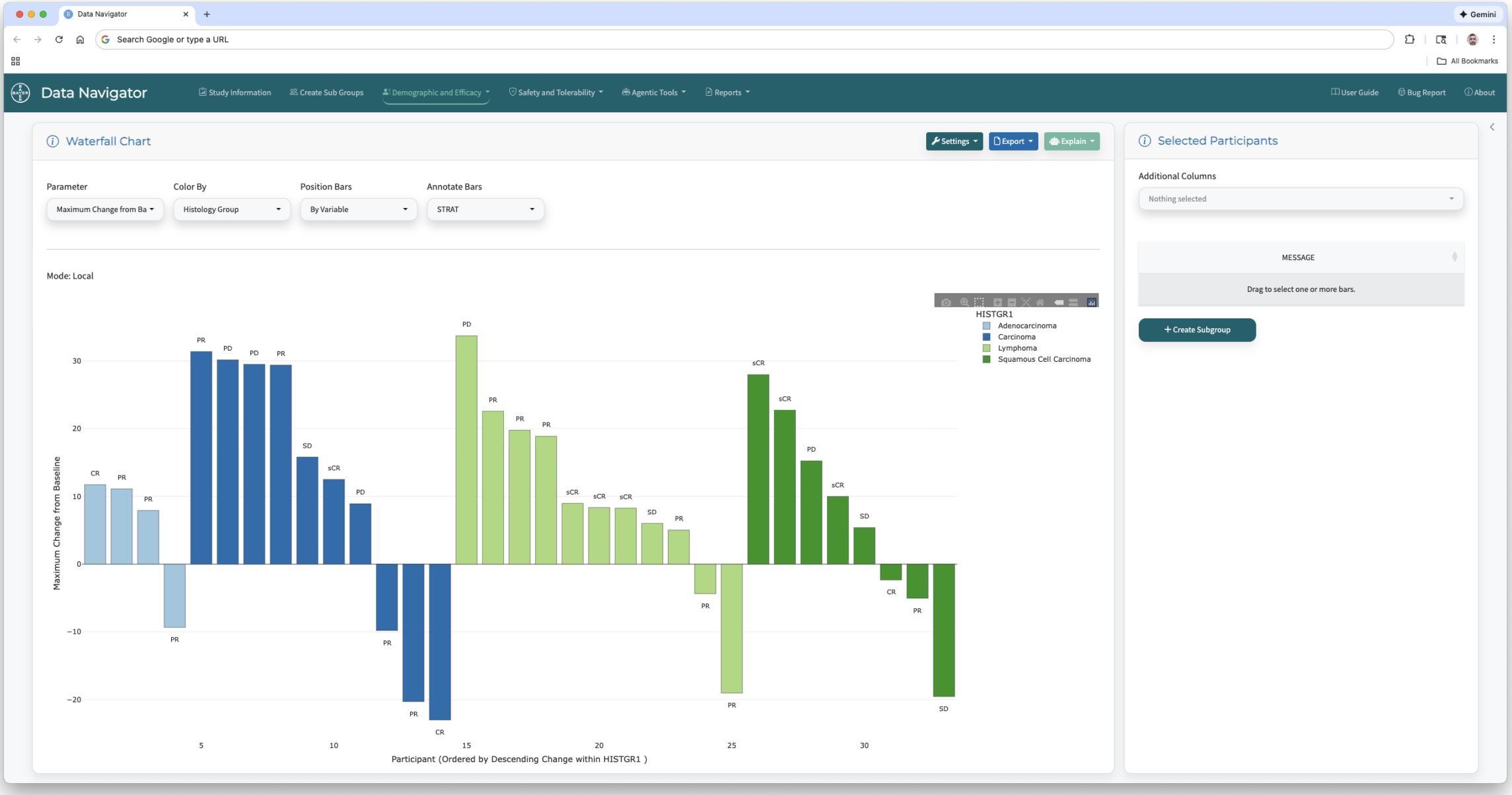
bayer.com

Back Up Slides


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
Demographic and Efficacy
Safety and Tolerability
Agentic Tools
Reports
User Guide
Bug Report
About
Box Chart
Settings
Export
Explain
Response Variable
Primary Variable
Secondary Variable
BMIBL
Biomarker Group 1
Race
Age AGE
WEIGHTBL WEIGHTBL
HEIGHTBL HEIGHTBL
BMIBL BMIBL
Subgroup_var Subgroup_var
Box Plot of BMIBL by BMCATEGORY1 colored by RACE
Drag to select points
RACE
AMERICAN INDIAN OR ALASKA NATIVE
ASIAN
BLACK OR AFRICAN AMERICAN
WHITE
35
30
25
20
15
BMIBL
Biomarker 1 Negative
Biomarker 1 Positive
BMCATEGORY1
Selected Points
Additional Columns
Smoking History (N)
USUBJID
SMOKHXN
01-701-1442
FORMER
01-703-1197
NEVER
Showing 1 to 2 of 2 entries
Previous
1
Next
+ Create Subgroup


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
Demographic and Efficacy
Safety and Tolerability
Agentic Tools
Reports
User Guide
Bug Report
About
Data Space
Search
USUBJID AGE SEX RACE ETHNIC ACTARM AGEGR2N BMCATEGORY1 HEIGHTBL WEIGHTBL
01-701-1015 63 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Placebo <65 Biomarker 1 Negative 58 120 25.0772885
01-701-1023 64 M AMERICAN INDIAN OR ALASKA NATIVE NOT HISPANIC OR LATINO Placebo <65 Biomarker 1 Positive 64 177 30.3786621
01-701-1028 71 M BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Positive 70 219 31.4197959
01-701-1033 74 M AMERICAN INDIAN OR ALASKA NATIVE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 69 195 28.7933207
01-701-1034 77 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Positive 61 138 26.0720236
01-701-1047 85 F WHITE NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Positive 58.5 148 30.4022207
01-701-1097 68 M WHITE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 66.5 172 27.3426423
01-701-1111 81 F WHITE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 62.2992125984252 132 23.9091502
01-701-1115 84 M AMERICAN INDIAN OR ALASKA NATIVE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 71.5 174 23.9272336
01-701-1118 52 M BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Placebo <65 Biomarker 1 Positive 71 157 21.8946637
01-701-1130 84 M ASIAN NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Negative 66.7992125984252 175 27.5708837
01-701-1133 81 F WHITE NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Negative 64 164 28.147460
01-701-1146 75 F WHITE NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Negative 58 119 24.8683115
Hello! I'm your Data Explorer Assistant.
I'm here to help you explore and analyze your clinical trial data. I can filter data, answer questions, and help you understand your dataset.
Try asking me:
Show me all subjects in the Placebo treatment arm
Filter the data where age is greater than 65 and sex is Female
What are the top 10 records sorted by baseline BMI?
Count how many subjects are in each treatment group
I'll display the results in the Data Space. What would you like to explore?
what is the mean age across all participants?
SELECT AVG(AGE) AS mean_age FROM raw_data
mean_age
75.09
The mean age across all participants is approximately 75.1 years. I determined this by calculating the average value of the AGE column for all entries in the raw_data table using SQL.
Enter a message...


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
Demographic and Efficacy
Safety and Tolerability
Agentic Tools
Reports
User Guide
Bug Report
About
Data Space
Search
USUBJID AGE SEX RACE ETHNIC ACTARM AGEGR2N BMCATEGORY1 HEIGHTBL WEIGHTBL
01-701-1034 77 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Positive 61 138
01-701-1047 85 F WHITE NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Positive 58.5 148
01-701-1111 81 F WHITE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 62.2992125984252 132
01-701-1115 84 M AMERICAN INDIAN OR ALASKA NATIVE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 71.5 174
01-701-1130 84 M ASIAN NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Negative 66.7992125984252 175
01-701-1133 81 F WHITE NOT HISPANIC OR LATINO Xanomeline High Dose ≥65 Biomarker 1 Negative 64 164
01-701-1153 79 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Positive 61.5 128
01-701-1181 79 F AMERICAN INDIAN OR ALASKA NATIVE NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Positive 61 147
01-701-1192 80 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Negative 59.5 140
01-701-1203 81 F BLACK OR AFRICAN AMERICAN NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Negative 63 146
01-701-1211 76 F ASIAN NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Positive 63 100
01-701-1324 79 M ASIAN NOT HISPANIC OR LATINO Xanomeline Low Dose ≥65 Biomarker 1 Positive 65.5 170
01-701-1363 81 F WHITE NOT HISPANIC OR LATINO Placebo ≥65 Biomarker 1 Positive 63.5 143
Show me all subjects in the Placebo treatment arm
Filter the data where age is greater than 65 and sex is Female
What are the top 10 records sorted by baseline BMI?
Count how many subjects are in each treatment group
I'll display the results in the Data Space. What would you like to explore?
what is the mean age across all participants?
SELECT AVG(AGE) AS mean_age FROM raw_data
mean_age
75.09
The mean age across all participants is approximately 75.1 years. I determined this by calculating the average value of the AGE column for all entries in the raw_data table using SQL.
filter participants above this value
-- Select all columns for participants older than the mean age
SELECT * FROM raw_data
WHERE AGE > (SELECT AVG(AGE) FROM raw_data)
The dashboard is now filtered to show only participants whose age is above the mean age of 75.1 years.
Enter a message...


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
Demographic and Efficacy
Safety and Tolerability
Agentic Tools
Reports
User Guide
Bug Report
About
Waterfall Chart
Settings
Export
Explain
Parameter
Maximum Change from Ba
Color By
Histology Group
Position Bars
By Variable
Annotate Bars
STRAT
Mode: Local
HISTGR1
Adenocarcinoma
Carcinoma
Lymphoma
Squamous Cell Carcinoma
Maximum Change from Baseline
Participant (Ordered by Descending Change within HISTGR1)
Selected Participants
Additional Columns
Nothing selected
MESSAGE
Drag to select one or more bars.
+ Create Subgroup


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
User Guide
Bug Report
About
Waterfall Chart
Parameter
Maximum Change from Ba
Color By
Histology Group
Position Bars
By Variable
Mode: Local
Maximum Change from Baseline
Participant (Ordered by Descending Change within HISTGR1)
Selected Participants
Additional Columns
Nothing selected
MESSAGE
Drag to select one or more bars.
+ Create Subgroup
This waterfall plot visualizes the maximum change from baseline in a measurement (likely tumor size or similar outcome) for each participant, grouped and color-coded by histological subtype (HISTGR1: Adenocarcinoma, Carcinoma, Lymphoma, Squamous Cell Carcinoma).
Key observations:
Each participant is represented by a bar, and bars are ordered by descending change within each tumor subtype.
Positive values (bars extending upward) indicate an increase from baseline, while negative values (downward bars) reflect a decrease.
Most participants in each group seem to exhibit positive maximum changes (i.e., increases), although there are notable exceptions, particularly among Carcinoma and Squamous Cell Carcinoma subtypes, where some individuals show marked decreases.
Response labels (e.g., CR = Complete Response, PR = Partial Response, SD = Stable Disease, PD = Progressive Disease, sCR = stringent Complete Response) are provided atop bars, indicating clinical response classification per participant.
Some patients with large negative changes (substantial decreases) are labeled as CR or PR, consistent with expected clinical response definitions.
Cautions and uncertainty:
The sample size for each group appears limited, so generalizations should be made cautiously.
The plot may reflect filtered data; thus, the totals and proportions for each tumor type may not represent the full dataset.
There is considerable variability within each subtype, indicating heterogeneity in individual responses.
Without statistical analysis, it's difficult to determine whether any apparent differences between subtypes are meaningful or significant.
Summary: There is intra- and inter-group variability in response as measured by maximum change from baseline, with some individuals across all histological subtypes experiencing marked decreases (potential clinical responses), while others experience increases. Further statistical analysis would be required to draw firm conclusions about group differences or associations with response.
Enter a message...

Data Navigator

Search Google or type a URL

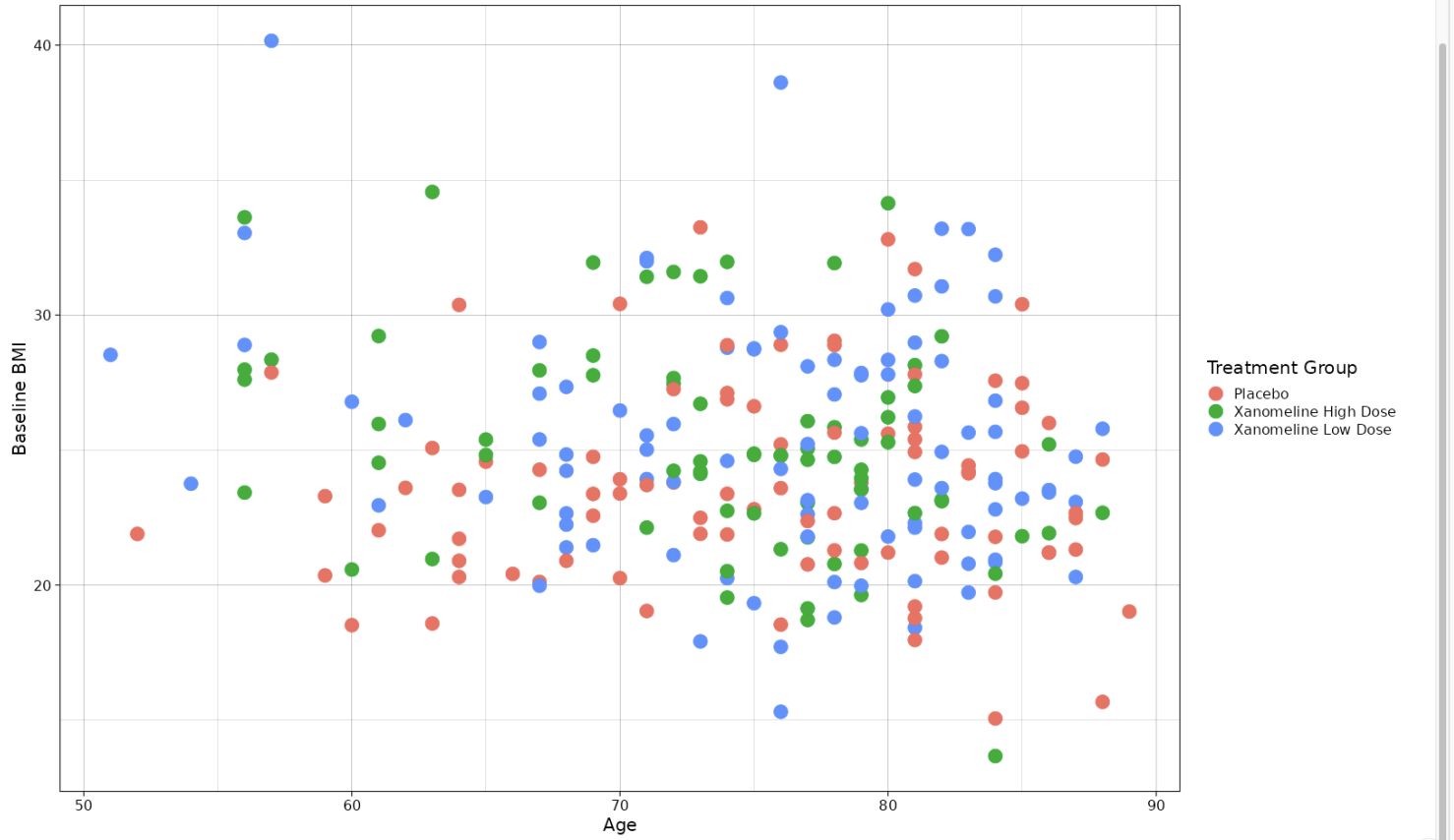
Gemini

All Bookmarks

Data Navigator

Study Information | Create Sub Groups | Demographic and Efficacy | Safety and Tolerability | Agentic Tools | Reports | User Guide | Bug Report | About

Plotting Space



GENERATED WITH AI

📊 Hello! I'm your Visualization Assistant (Grapher).

I'm here to help you create graphs and visualizations from your clinical trial data. I can generate charts, plots, and custom visualizations based on your requests.

Try asking me:

- Create a scatter plot of age versus baseline BMI, colored by treatment group ✦
- Make a box plot showing the distribution of response values by treatment arm ✦
- Generate a bar chart of subject counts by sex and race ✦
- Create a histogram of baseline weight with different colors for each treatment ✦

I'll display your visualizations in the Plotting Space. What visualization would you like to create? 🎨

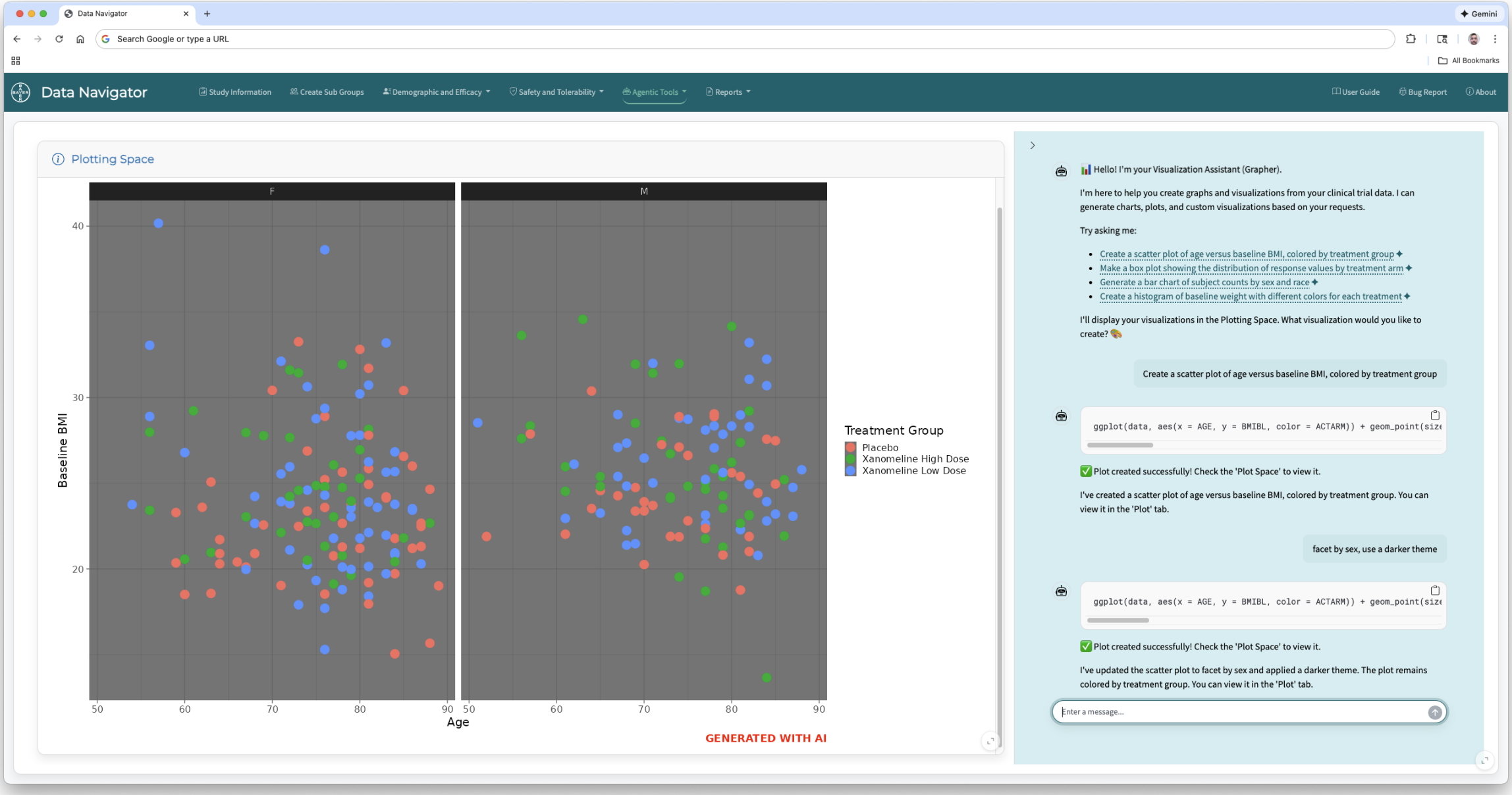
Create a scatter plot of age versus baseline BMI, colored by treatment group

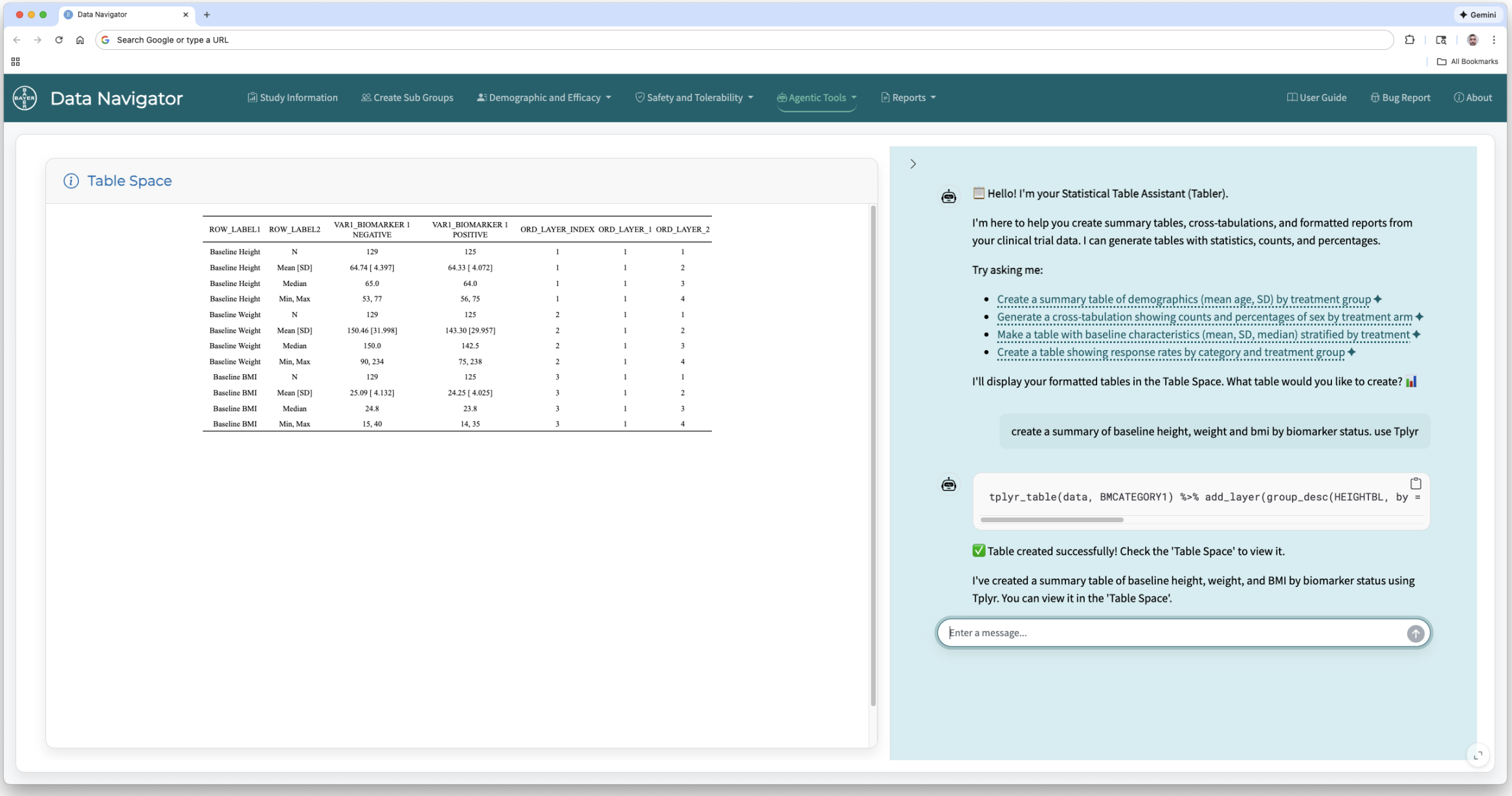
```
ggplot(data, aes(x = AGE, y = BMIBL, color = ACTARM)) + geom_point(size
```

✅ Plot created successfully! Check the 'Plot Space' to view it.

I've created a scatter plot of age versus baseline BMI, colored by treatment group. You can view it in the 'Plot' tab.

Enter a message...


Data Navigator
Search Google or type a URL
Gemini
All Bookmarks
Data Navigator
Study Information
Create Sub Groups
Demographic and Efficacy
Safety and Tolerability
Agentic Tools
Reports
User Guide
Bug Report
About
Plotting Space
F
M
Baseline BMI
40
30
20
50
60
70
80
90
Age
Treatment Group
Placebo
Xanomeline High Dose
Xanomeline Low Dose
GENERATED WITH AI
Hello! I'm your Visualization Assistant (Grapher).
I'm here to help you create graphs and visualizations from your clinical trial data. I can generate charts, plots, and custom visualizations based on your requests.
Try asking me:
Create a scatter plot of age versus baseline BMI, colored by treatment group
Make a box plot showing the distribution of response values by treatment arm
Generate a bar chart of subject counts by sex and race
Create a histogram of baseline weight with different colors for each treatment
I'll display your visualizations in the Plotting Space. What visualization would you like to create?
Create a scatter plot of age versus baseline BMI, colored by treatment group
ggplot(data, aes(x = AGE, y = BMIBL, color = ACTARM)) + geom_point(size
Plot created successfully! Check the 'Plot Space' to view it.
I've created a scatter plot of age versus baseline BMI, colored by treatment group. You can view it in the 'Plot' tab.
facet by sex, use a darker theme
ggplot(data, aes(x = AGE, y = BMIBL, color = ACTARM)) + geom_point(size
Plot created successfully! Check the 'Plot Space' to view it.
I've updated the scatter plot to facet by sex and applied a darker theme. The plot remains colored by treatment group. You can view it in the 'Plot' tab.
Enter a message...



ROW_LABEL1	ROW_LABEL2	VAR1_BIOMARKER 1 NEGATIVE	VAR1_BIOMARKER 1 POSITIVE	ORD_LAYER_INDEX	ORD_LAYER_1	ORD_LAYER_2
Baseline Height	N	129	125	1	1	1
Baseline Height	Mean [SD]	64.74 [4.397]	64.33 [4.072]	1	1	2
Baseline Height	Median	65.0	64.0	1	1	3
Baseline Height	Min, Max	53, 77	56, 75	1	1	4
Baseline Weight	N	129	125	2	1	1
Baseline Weight	Mean [SD]	150.46 [31.998]	143.30 [29.957]	2	1	2
Baseline Weight	Median	150.0	142.5	2	1	3
Baseline Weight	Min, Max	90, 234	75, 238	2	1	4
Baseline BMI	N	129	125	3	1	1
Baseline BMI	Mean [SD]	25.09 [4.132]	24.25 [4.025]	3	1	2
Baseline BMI	Median	24.8	23.8	3	1	3
Baseline BMI	Min, Max	15, 40	14, 35	3	1	4