# A BIRD'S-EYE VIEW OF CLINICAL TRIALS

## IMT 562 - MAKEOVER MONDAY

Jwala Mukhi Suresh

# ABOUT THE DATASET

This dataset, A bird's-eye view of clinical trials is from Makeover Monday 2019 week 33. The original data source is AERO Data Lab. The data analyzes all the registered clinical trials from 10 large pharmaceutical companies — AbbVie, Bayer, Gilead, GSK, Johnson & Johnson, Merck, Novartis, Pfizer, Roche, and Sanofi from 1984 to 2019.

Link to the Makeover Monday visualization: https://data.world/makeovermonday/2019w33

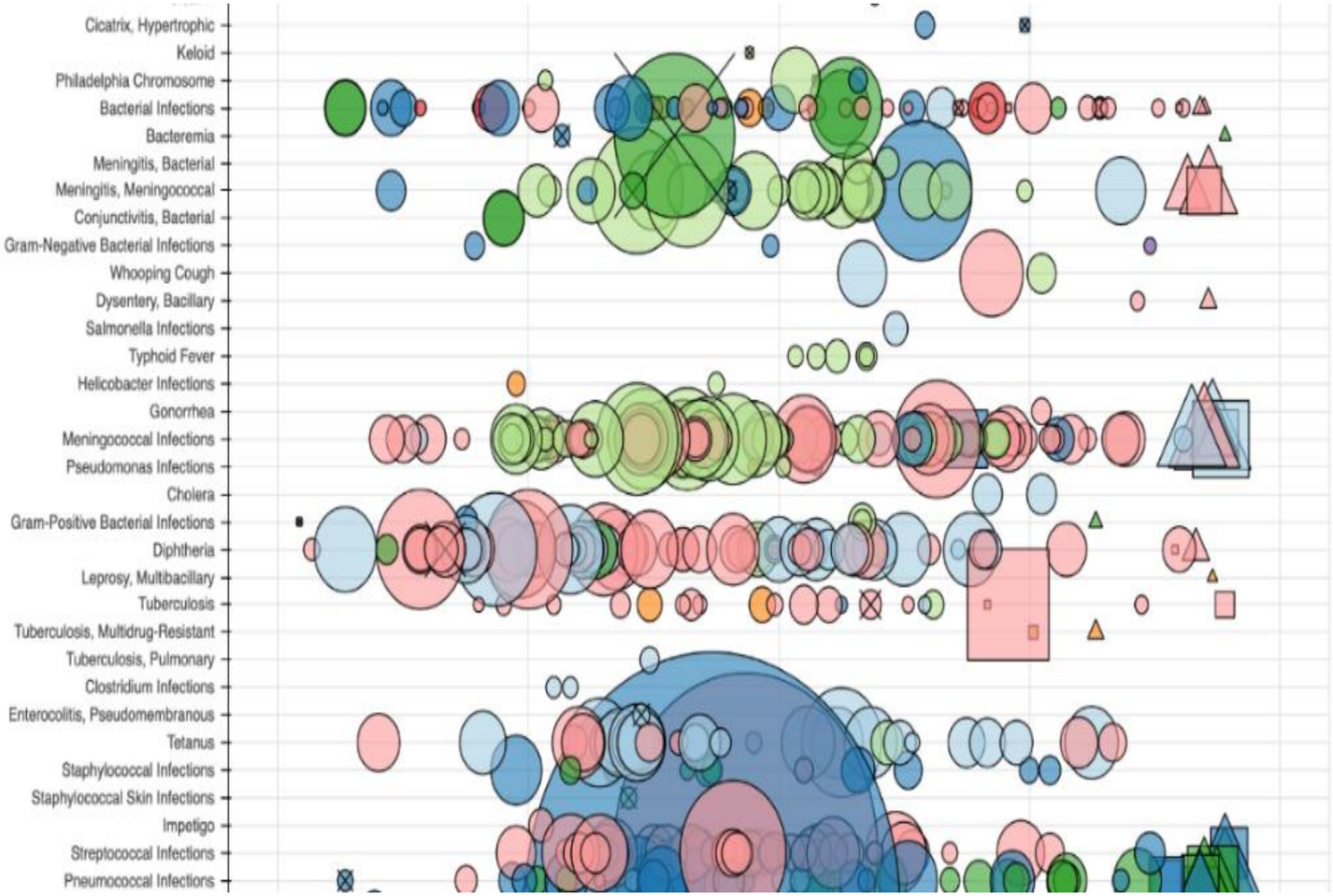
Link to original source: https://www.aerodatalab.org/birds-eye-view-of-research-landscape

Total number of records: **13,748**

| Columns | Datatype | Data Description |
|---|---|---|
| **NCT** | String | Unique id for the trial |
| **Sponsor** | String | Name of sponsor (10 sponsors) |
| **Title** | String | Title of the trial |
| **Summary** | String | Summary of the trial |
| **Start_Year** | Year | Trial starting year (1984 – 2020) |
| **Start_Month** | Integer | Trial starting month |
| **Phase** | String | Phase of the trial |
| **Enrollment** | Integer | Number of people enrolled in the trial |
| **Status** | String | Status of the trial |
| **Condition** | String | The health condition on which the clinical trial is conducted |

# INITIAL VISUALIZATION

Every node corresponds to a registered trial from one of the 10 companies. Trials are then organized from oldest to newest on the x-axis and by patient population or disease under study on the y-axis. The color of a node indicates the company. The shape indicates the trial's status — completed studies are circles, recruiting studies are triangles, terminated studies have an "x" through them — and the size represents the number of patients enrolled. (Skerrett, 2019)



# POSITIVE ASPECTS OF THE VISUALIZATION

- From a simple glance, the chart reveals the incredible volume of research that has been undertaken.
- A single chart is used to represent all details about the clinical trials that took place from 1984 to 2019. Users can get all details just by zooming in or hovering over the bubbles without any additional filters.
- The tooltip includes all details about a particular trial and when user clicks on the bubble, they are redirected to the corresponding trial's website.
- The size and concentration of bubbles helps to understand which the active or less active areas of experiment are based on the concentration of bubbles in an area.

# CRITIQUE OF THE VISUALIZATION

- ***Design:*** The initial visualization is very complex and overcrowded by trying to include every detail of over 13000 trials in a single chart without any filters. Since most of the bubbles are overlapping in the chart the colors, size and shape are not helpful to distinguish between the different trials and is making the data incomprehensible. The image added here is only a part of the actual visualization as the complete one is very long.
- ***Position:*** The visualization had tried to include all condition/disease mentioned in the dataset along the y axis, which makes the chart too lengthy and reduces the readability.
- ***Colors:*** The use of too many colors without any legend to explain what it represents makes the visualization confusing for the user.
- ***Size and Shape:*** The visualization is making use of bubbles with different size and shapes but fails to explain what each of them represents. Smaller bubbles and the shapes are hard to notice as most of the data in visualization is crowded.

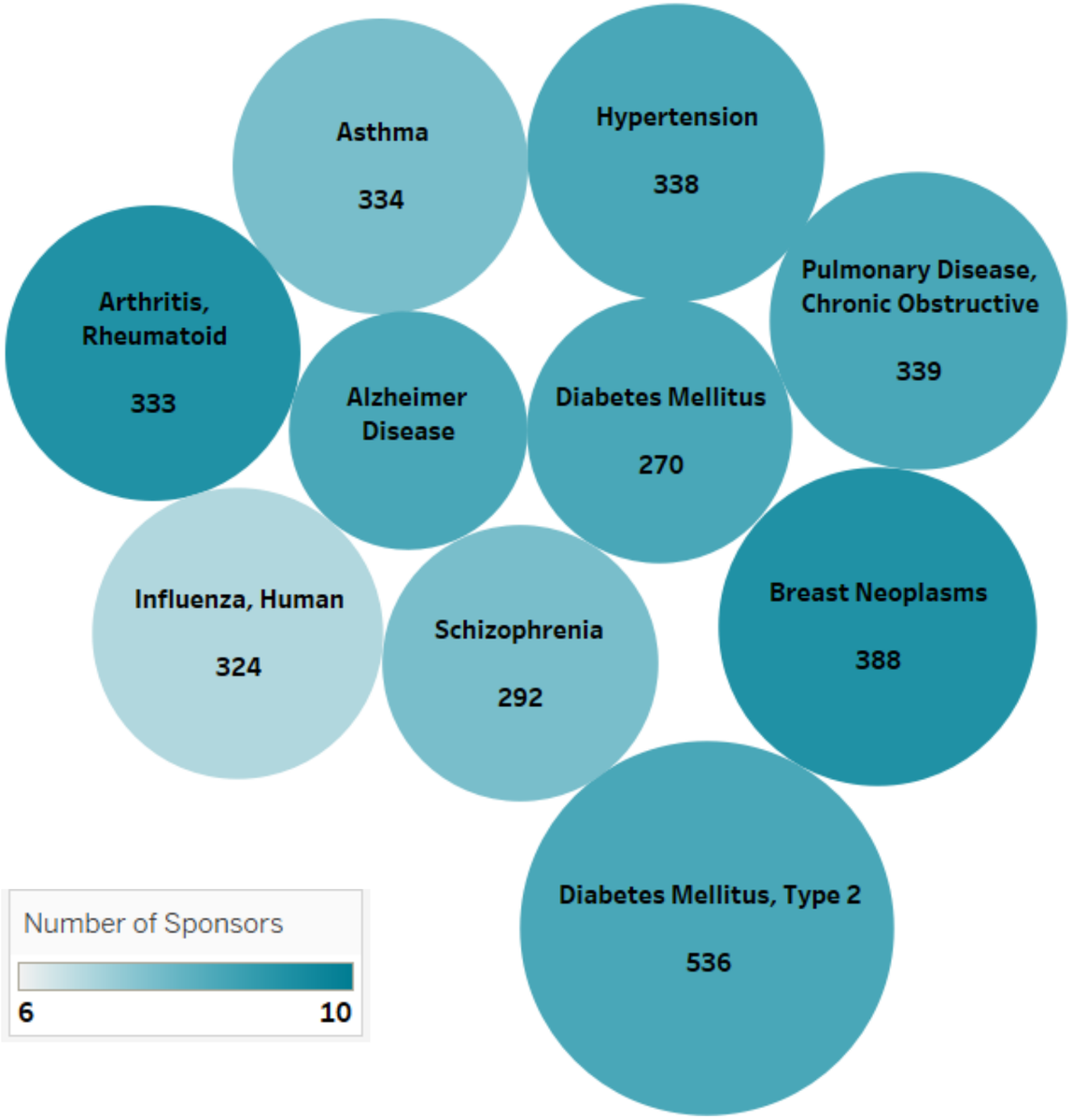
# PRIORITY LIST OF IMPROVEMENTS

- ***Improve design and reduce complexity:*** Make the visualization simple by using multiple charts, so that, the user can concentrate on separate details without feeling overwhelmed.
- ***Add legends:*** To help users decode the visual representation easily.
- ***Add filters:*** To limit and customize the data displayed on visualization in a dashboard.
- Multiple visualizations to highlight the important details from the chart
- Use a single theme for the entire visualization

# EXPLORATORY DATA ANALYSIS QUESTIONS


- Which were the top 10 conditions for which most trials were done for?
- How many trials did each sponsor conduct and how many participants?
- How did the research trend for each sponsor change over the years?
- How many projects had suspended, terminated, withdrawn, unknown status for each sponsor?
- Which were the top 10 conditions based on total number of enrollments?
- Which are the top 10 conditions that had the greatest number of shortest (completed in either Early Phase 1/Phase 1)/longest trials (completed in Phase 4)?

# TOP 10 CONDITIONS WITH MOST TRIALS

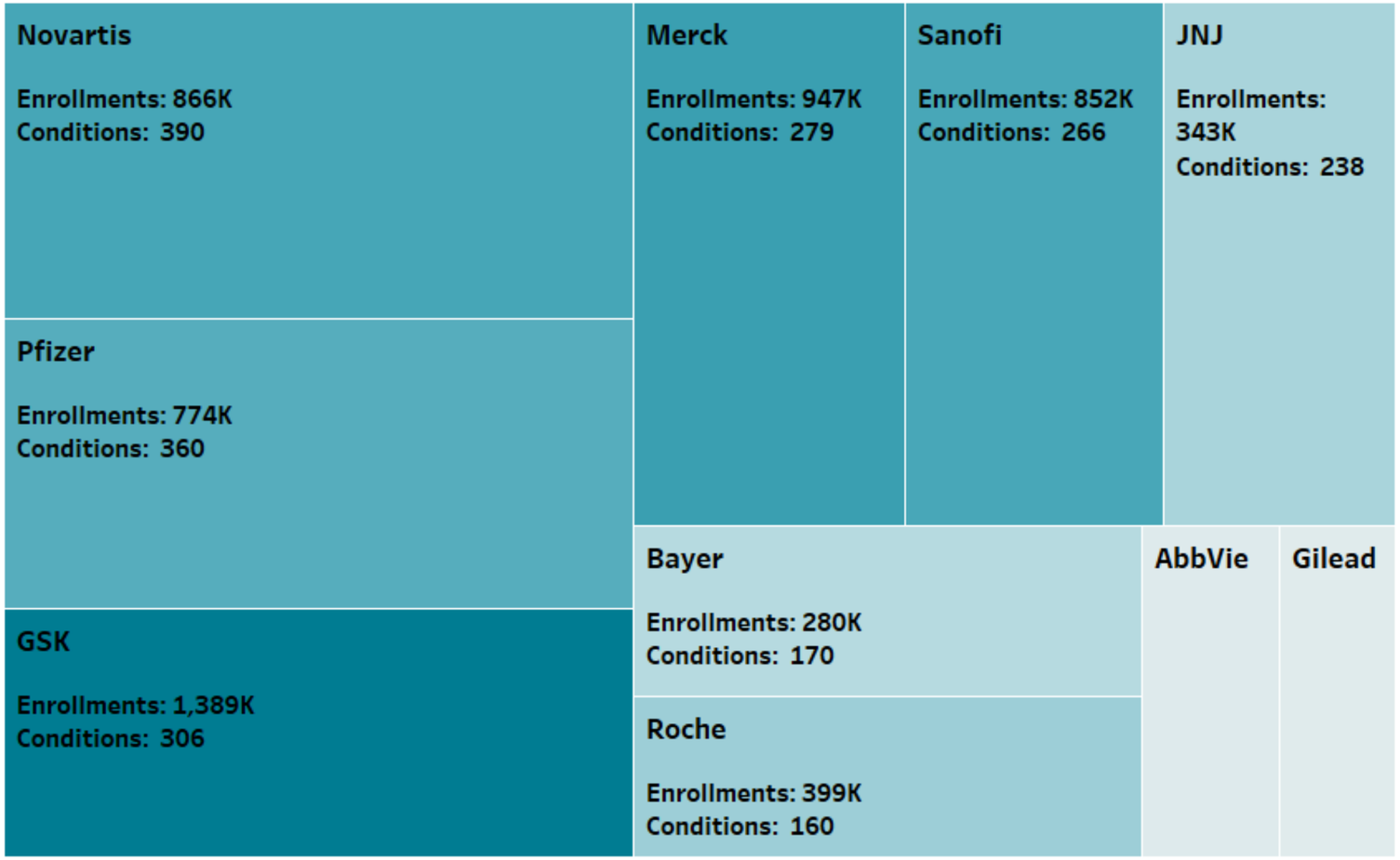


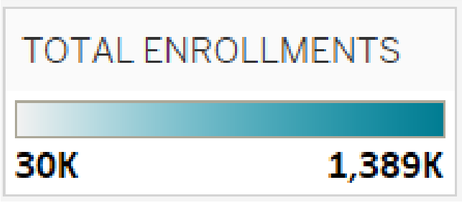

- The bubble chart shows the top 10 conditions for which most trials have been conducted. The size of the bubble shows the number of trials performed for a condition and hue shows the number of distinct sponsors who have conducted trials for the condition.

**Rationale for this visualization**

- From the huge list of trials from 1984 to 2019, this bubble chart helps to easily identify the prominent 10 conditions that sponsors focused for trials which was hard to understand from the original visualization.

# TOTAL CONDITIONS & TOTAL ENROLLMENTS FOR EACH SPONSOR



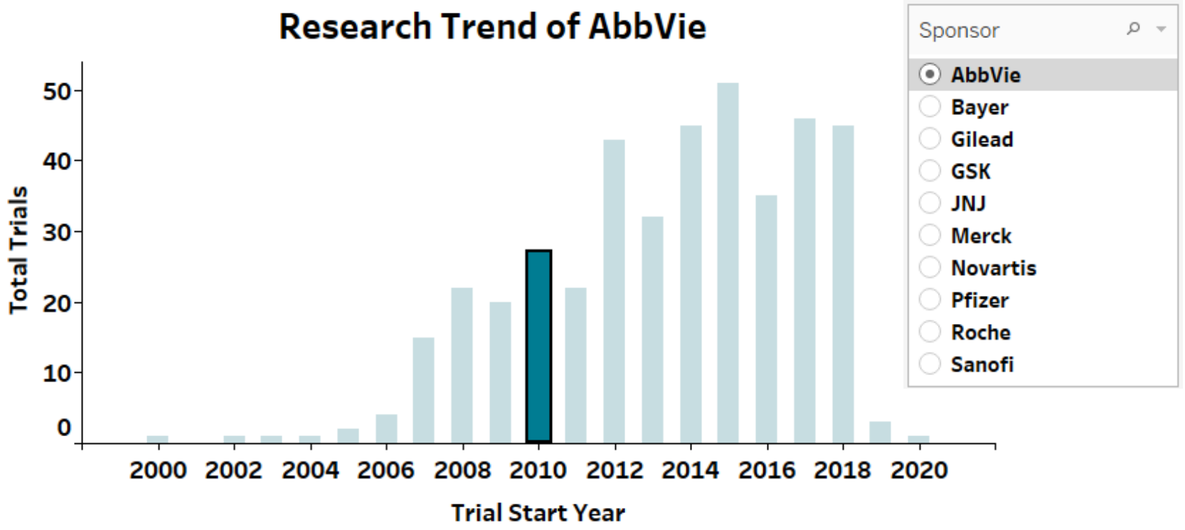



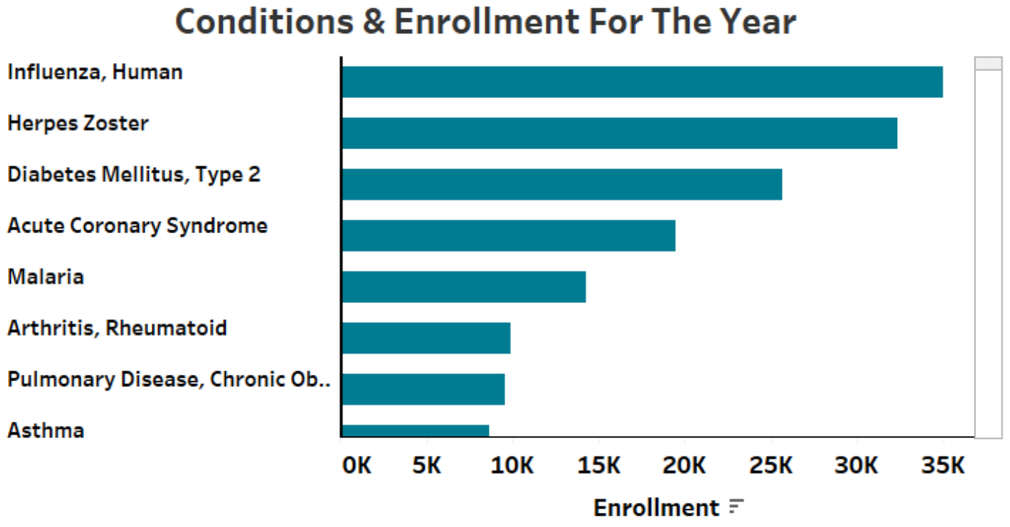

- Size is used to show the total conditions for which the trials were conducted by a particular sponsor and hue indicates the total enrollments they had for their trials. All these details are also included in the tooltip.

**Rationale for this visualization**

- The use of a treemap helps to understand the multiple dimension in a single chart.
- This chart gives the high-level overview on sponsors.

# RESEARCH TREND OF EACH SPONSOR OVER THE YEARS



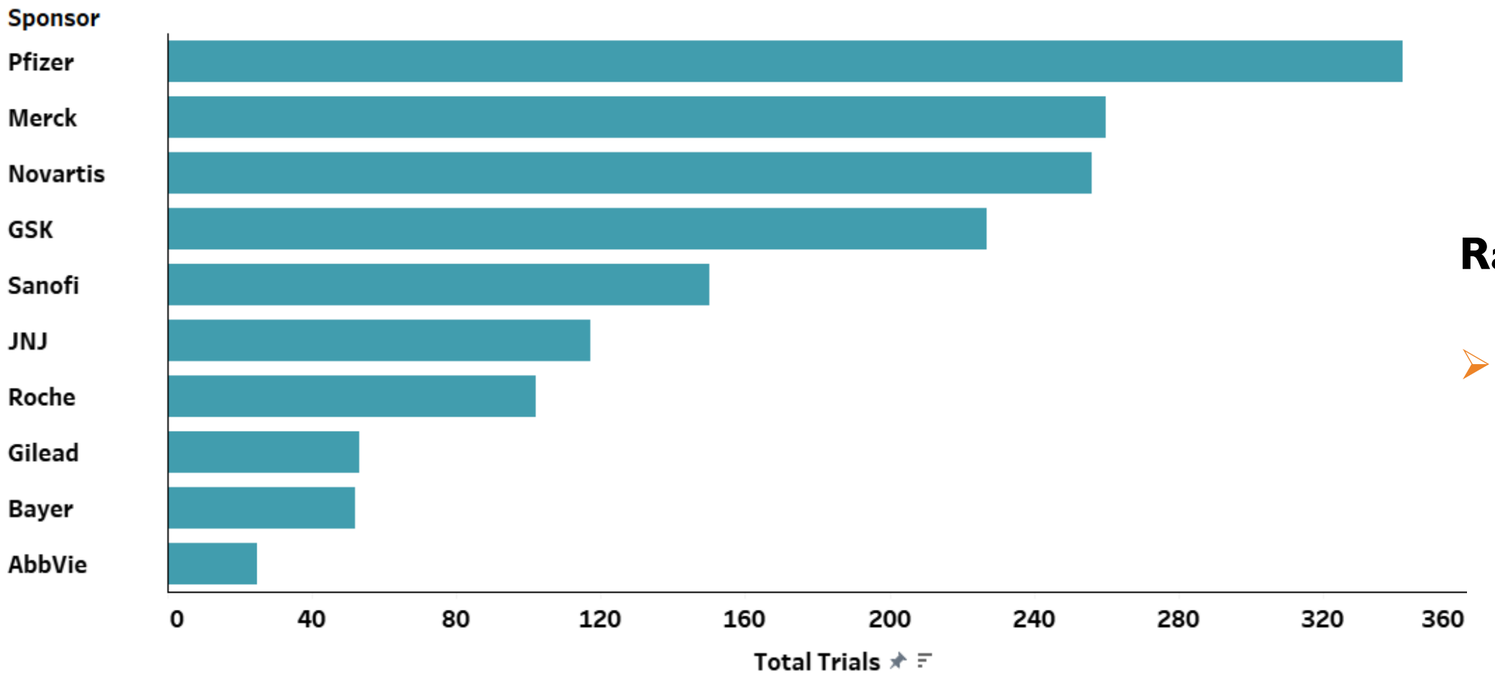




- The two bar charts together shows the research trend of a sponsor throughout the years. On choosing the sponsor the first chart shows the total trials performed over the years and on clicking the bar for a year the lower graph filters and shows the conditions & total enrollments for the year.

**Rationale for this visualization**

- Using two separate bar charts and the use of interactivity makes it easy for the users to understand the research pattern of different sponsors and how active they were in conducting researches for different conditions and how many participants they enrolled in each trial.

# TOTAL TRIALS THAT WERE SUSPENDED, TERMINATED, WITHDRAWN, UNKNOWN STATUS FOR EACH SPONSOR

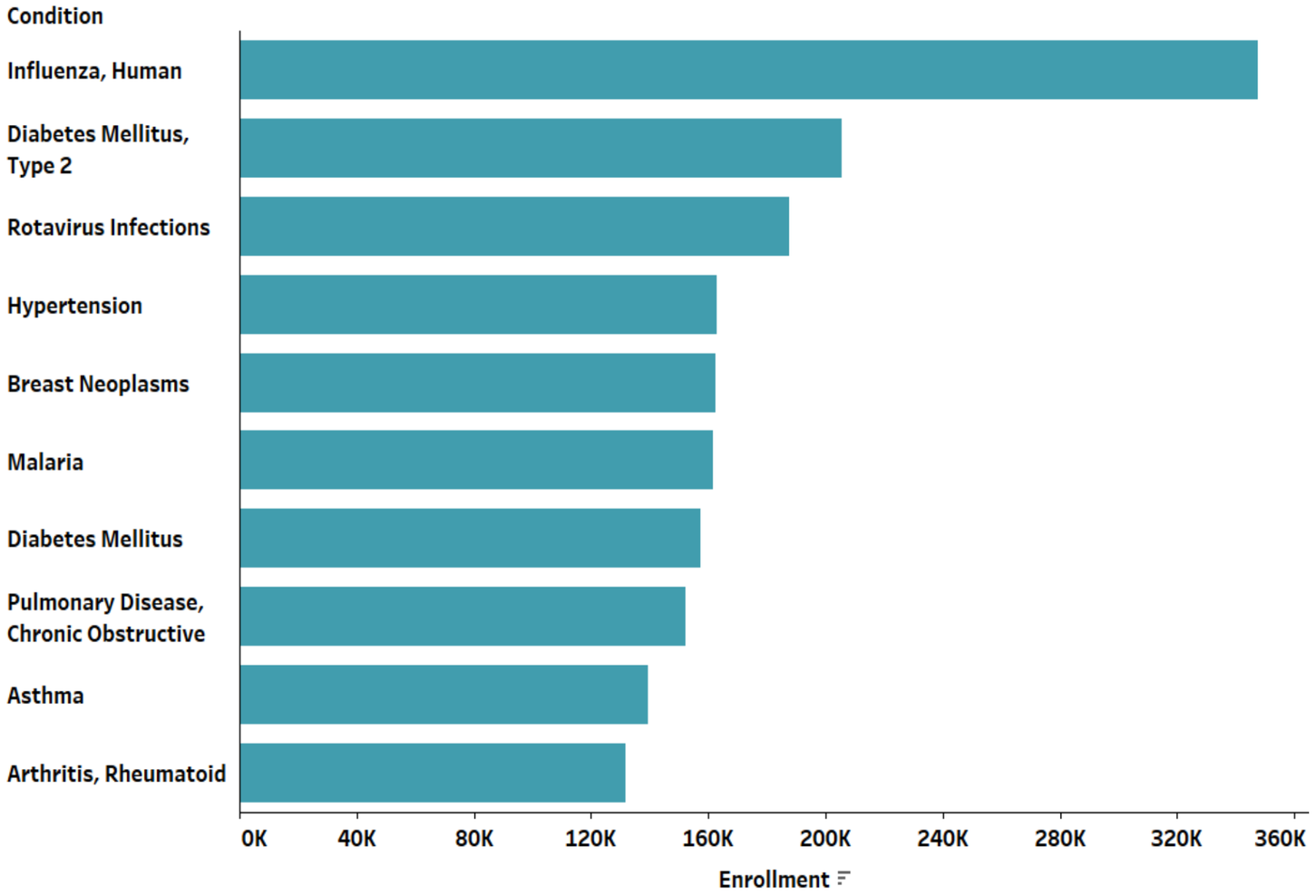



- The visualization shows the total trials by each sponsor that were suspended, terminated, withdrawn, and unknown status.

**Rationale for this visualization**

- The initial visualization had used multiple representations for showing the projects which cannot be understood without zooming in.
- The use of a simple bar chart helps to understand the total trials conducted by the sponsors that weren't successfully completed, in a single glance.

# TOP 10 CONDITIONS BASED ON TOTAL ENROLLMENTS

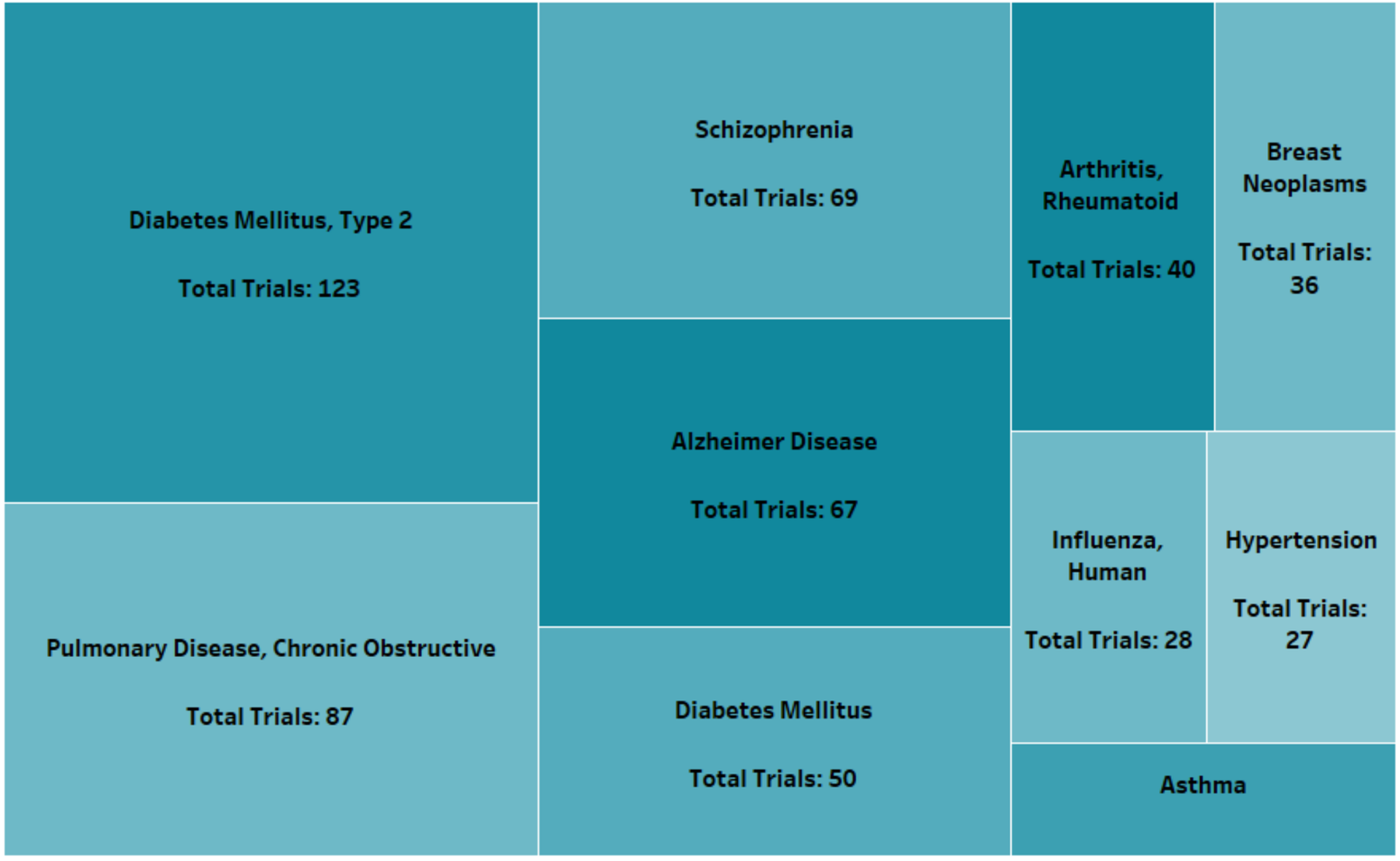


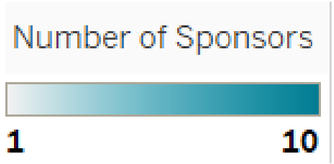

The bar chart shows which were the top 10 conditions for which most people enrolled.

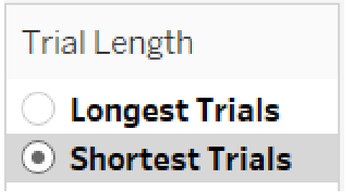
**Rationale for this visualization**

- The initial visualization showed only enrollments per trial, but it didn't help to understand which conditions had the most enrollments.
- Using a bar chart makes it easy for the user to easily identify the conditions which had the most enrollments over the entire period from 1984 to 2019

# TOP 10 CONDITIONS WITH SHORTEST/LONGEST TRIALS




- The visualization uses filter to show the trials that completed in Early Phase 1/Phase1 as shortest trials and the trials that completed in Phase 4 as the longest over the period of 1984-2019. The box size represents the count of trials and hue the count of sponsors.

**Rationale for this visualization**

- The initial visualization didn't classify trials based on the Phases and included these details only in the tooltip which was hard to follow and compare trials lengths for different conditions. This chart helps in easily identifying the top 10 conditions where trials completed early or later.

# SUMMARY & KEY INSIGHTS

- Over the period of 26 years from 1984-2019, there were 13748 trials conducted by 10 sponsors on a total of 867 distinct conditions with a total enrollment of 6.06M.
- "Diabetes Mellitus, Type 2" was the condition with the greatest number of trials done over the course. There were 536 trials done and 9 out of 10 sponsors were involved along with close to 205k enrolments.
- Novartis was the sponsor who researched on the greatest number of conditions. Novartis conducted trails on 396 unique conditions with close to 855k enrolments., followed by Pfizer with 360 conditions and 774k enrolments.
- Condition "Human Influenza" had the greatest number of enrolments (~348k) and interestingly, the second in the line to condition "Diabetes Mellitus, Type 2" had only 205k.

# REFERENCE

Skerrett, P. (2019, July 17). *A bird's-eye view of clinical trials provides new perspectives on drug research and development*. STAT. https://www.statnews.com/2019/07/18/clinical-trials-birds-eye-view-drug-development/

# TABLEAU PUBLIC LINKS

DASHBOARD: Clinical Trials (1984-2019)

PACKAGED WORKBOOK:

Clinical Trials Packaged Workbook