**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 2020.

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**

Chart, bubble chart

Description automatically generated

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 2020. 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.

**Exploratory Data Analysis Questions**

* Which were the conditions (top 10) for which most trials were done for?
* How many trials did each sponsor’s conduct and how many participants?
* How did the research activity in each condition change over the year?
* What is the total number of suspended, terminated, withdrawn, unknown status projects and which sponsor had the greatest number?
* Which were the top 10 trials with the greatest number of enrollments?
* Total count of patients
* Smallest trials that were completed

1987 trials completed in either Early phase1/Phase 1

* Largest trials

1659 trials completed in Phase 4

* It is also interesting to note the gaps in activity. What happened in migraine research after 2010? Why have these 10 companies initiated so few migraine trials since then? By contrast, the activity in Alzheimer’s research looks steadier — although it has more terminated trials.

Even though trial termination is not necessarily a bad thing — some trials are terminated early because the experimental drug is found to be highly effective — trials are not designed with the intention of being stopped early. So, a cluster of terminations in a particular area demands an explanation.

**Summary & Key Insights**

**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/>