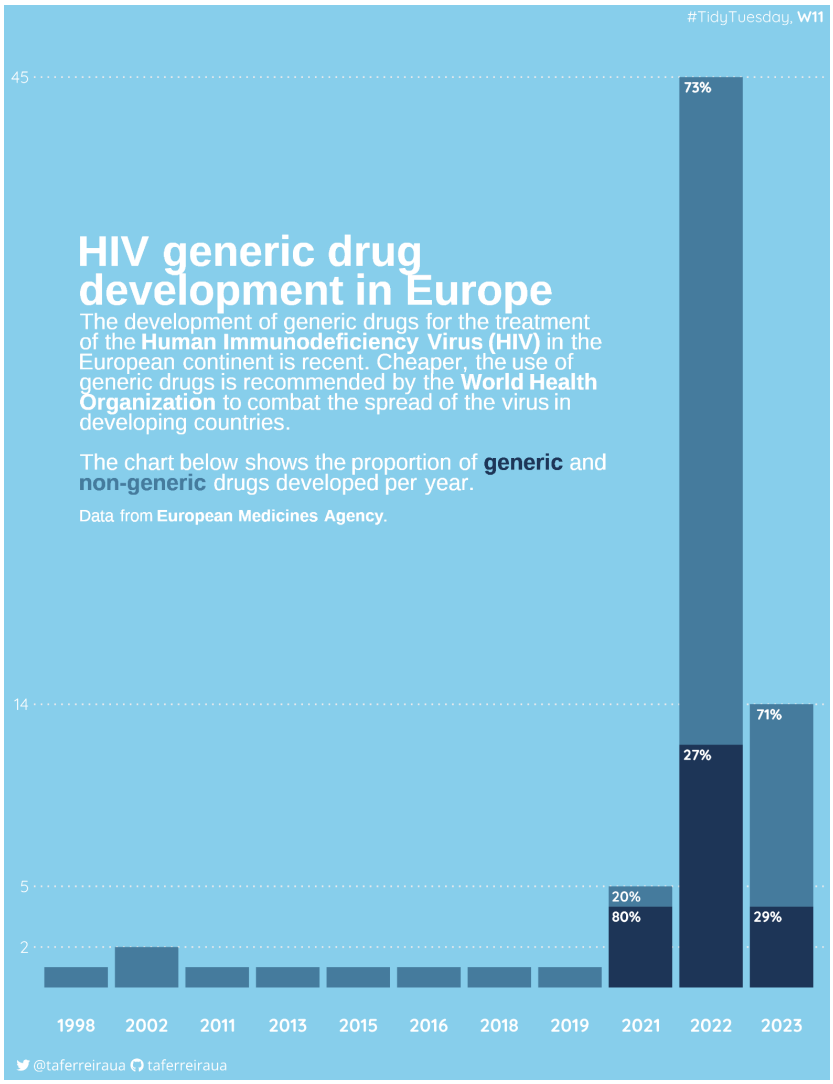
HW4 Report

Critiqued Visualization:
https://twitter.com/taferreiraua/status/1636056091967594502/photo/1
Source data:
https://github.com/rfordatascience/tidytuesday/blob/master/data/2023/2023-03-14/drugs.csv

Existing Visualization Critique



In the visualization that we chose to critique, the author made a bar chart depicting the development of drugs for the treatment of Human Immunodeficiency Virus, otherwise known as HIV, from the years 1998-2023. The author also split the bars according to whether the treatment was generic or non-generic (darker blue versus medium blue color). From our understanding, the

author was trying to show how the use of generic drugs has been recent (starting in the year 2021) and that they are perhaps a better alternative as recommended by the World Health Organization. There are many positive and negative aspects to this visualization, which we will explore now.

To begin with some positive aspects, we thought that the darker blue color stood out well against the lighter blue color. This contrast allows the user to easily see at what point generic drugs were beginning to be utilized. It is straightforward to see the trend of increased generic drug use in the years 2021-2023 which was arguably the author's main point from the visualization. Another positive aspect was the use of dotted white grid lines connecting to the bars. From the choice of both the scale and the use of these lines, it was painless to discern the height of each bar and the corresponding numerical value. Lastly, another positive aspect to the visualization was the use of percentages within the split bar chart. These percentages are helpful to the user because they show how much of each bar each colored region is actually taking up. With the longer bar representing 2022 especially, it can be challenging for the user to discern how much area each section represents; thus, the percentages are a great way to numerically display this information.

In terms of negative aspects and flaws of the visualization, we thought of many ways in which the visualization could ultimately be improved. To begin with, although the percentages shown in the rectangular bars were helpful in giving relative size information, they could also be misleading to the user. For example, in 2022, the light blue bar, or non-generic drugs, represent 73% of drugs made that year. Right next to it, the light blue bar of 2023 also contains 71% of drugs made so far in that year. However, the sizes of these bars differ greatly from one another because the percentages are of course relative to the total counts for each year. This is confusing to the user because they would naturally believe that the values of 71 and 73 should be similar in size to one another. Thus, we believe that it would have been best for the author to focus on either counts or percentages respectively. In our improved visualization, we will at first focus on the actual numerical counts of drugs created rather than percentages. However, once we go into further detail of a specific category, we can then break down the data into percentages of generic versus non generic as well as other distinctions that can be made from the data.

Another flaw of this visualization is the inconsistent time scale of the bar chart. The chart seems to jump from 1998 to 2002 to 2011 with no indication that the scale is taking large jumps through time. Perhaps the author created the scale in this method in order to make sure there were no bars with zero data points. However, alternatively, the author could have made a bin representing a range of years to further emphasize the point that HIV drug development has only recently really expanded. In the revised version of this visualization, we will make sure to include consistent time scales if we are to employ any. In addition, we will not make time the primary axis of our overview visualization. Instead, we wish to focus more on the different types of therapeutic areas that certain drugs are targeting, and how the number of drugs may differ across these types. Once

we zoom in on a specific type, however, we will explore the temporal component to the data as well, similar to the original visualization.

Yet another flaw of the visualization is the fact that it only focused on one specific target area for drugs: HIV. From the bar chart alone, we can notice a trend for the development of HIV drugs specifically; however, there is no contextual evidence in terms of other therapeutic areas. It would have been helpful to compare the generic versus non generic drug trend for other types of drugs as well in order to conclude if the trend can be generalized or not. In our improved visualization, we will compare various categories with HIV in order to make more broad conclusions, and compare across several subsets of data. We think that this broader use of data will give way to more interesting insights.

Lastly, we think the use of an all-blue color scale is most likely not the best choice. Although the darker blue does contrast against the more medium-colored blue, we believe that choosing colors with different hues would have been the smarter choice. The use of a color-blind friendly color scale would more effectively make the different bars pop out at the user immediately.

Ultimately, this visualization proved to be a good starting point for comparing the development of drugs over a time period of roughly 20 years. However, there are various ways in which we can expand this visualization to more intriguing directions. Namely, we will focus not just on HIV, but on the top ten more common therapeutic areas for drug development. That way, we can compare trends across multiple groups instead of focusing on just one. Moreover, we will begin with an overview that displays the counts of each drug group much like the histogram in the original visualization, but with the drug groups instead of time on the x axis. The user will then be able to click on any of the groups to uncover more detailed information, such as the breakdown of generic versus non generic drugs, approved vs withdrawn, and the time at which the drugs were developed. Lastly, we want the user to get a snapshot of each drug more specifically, rather than it just being represented pictorially on a histogram part of the rectangle. Therefore, we will implement some sort of interaction in which the user will get a tabular display of information about each specific drug including its scientific name, common name, as well as a description of the drug.

Improved Visualization

Revised Visualization: https://gregarious-empanada-d372f2.netlify.app/

For the revised version of our visualization, we wanted to present a more overviewed, detailed version of the preliminary visualization that we found to critique above. In its beginning primary layout, we created a simple bar chart much like the source inspiration. However, in our visualization, we decided to include the top 10 most common therapeutic areas from the dataset in order to allow the user to explore trends among multiple categories instead of just one. Next, we added on to the existing capabilities by including a hovering interaction for both filtering by approval status (withdrawn versus approved) as well as the counts for each bar in the bar graph. In this way, users are able to compare the counts for only approved or withdrawn if they so

chose, and they can also see exact numerical values for each bar in the chart. We believe that using just the actual count is a better way of representing data then the percentages used in the source inspiration that were often misleading.

Next, the user is able to click on a specific bar in order to be shown more detailed information on that therapeutic drug group. We have multiple interactions that aid the user. First, once the user clicks on a specific bar, the rest of the bars will have a lower opacity so that the bars selected stand out and the user is reminded of which bar they selected. Secondly, the page will scroll automatically down to the bottom of the page where the detailed information lies. In addition, the therapeutic group that is selected will be presented as a title at the top of the page. Our visualization includes a timeline that shows when the drugs, both approved and withdrawn, were first published. We hope that this timeline will give a better sense as to the scientific discovery of drugs in this specific area, and how this compares to both approved and withdrawn drugs as well as the other categories.

Lastly, the user is also able to hover over specific points in the timeline visualization in order to get more details about that specific drug. We believed that this would be a great addition to the source inspiration so that the user can explore the different drugs and not just the overarching, general trends. We included information such as its common name, the active ingredient, and a URL that the user can go to in order to learn even more about the drug if they so choose. Ultimately, we believe that our improved visualization goes above and beyond our source inspiration. Although the original visualization had potential, it lacked context in terms of other drugs and their corresponding therapeutic categories. With our improved visualization, we would like the users to be able to make comparisons across groups, as well as learn more about the different drugs that were approved by the FDA versus those that were withdrawn.