ethical viz

team

2024-02-24

```
# Load libraries
library(tidyverse)  # For ggplot, dplyr, and friends
```

```
# Load the original csv file
distributor <- read_csv("data/distributor.csv")

# View the content of the loaded CSV file
View(distributor)

# Display the dataframe
knitr::kable(head(distributor))
```

Reporter_family	total_dosage_unit	total_mme	total_records
McKesson Corporation	12970810	98012825	19060
Miami-Luken	12612410	75657518	8707
H. D. Smith	8967110	61367298	9794
AmerisourceBergen Drug	5198890	54812028	6987
Cardinal Health	4029300	26162297	6038
Quest Pharmaceuticals Inc	1099260	6215122	336

```
# Load the tidyverse package for data manipulation
library(tidyverse)

# Display the dataframe with formatted values
distributor_formatted <- distributor %>%
  mutate(across(where(is.numeric), scales::comma))
knitr::kable(head(distributor_formatted))
```

Reporter_family	total_dosage_unit	total_mme	total_records
McKesson Corporation	12,970,810	98,012,825	19,060
Miami-Luken	12,612,410	75,657,518	8,707
H. D. Smith	8,967,110	61,367,298	9,794
AmerisourceBergen Drug	5,198,890	54,812,028	6,987
Cardinal Health	4,029,300	26,162,297	6,038
Quest Pharmaceuticals Inc	1,099,260	6,215,122	336

Ethical Visualisation - Blue Team

The visualization presented aims to depict the distribution of total dosage units by the top three pharmaceutical distributors in Mingo County, West Virginia. It effectively communicates the magnitude of dosages each distributor has contributed, with values represented in millions. The bar chart clearly distinguishes between the distributors: McKesson Corporation, Miami-Luken, and H.D. Smith, using distinct colors to represent each one. The lengths of the bars make it easy to compare the distributors immediately, with McKesson Corporation appearing to have distributed the highest number of dosage units, closely followed by H.D. Smith, and Miami-Luken contributing the least of the three.

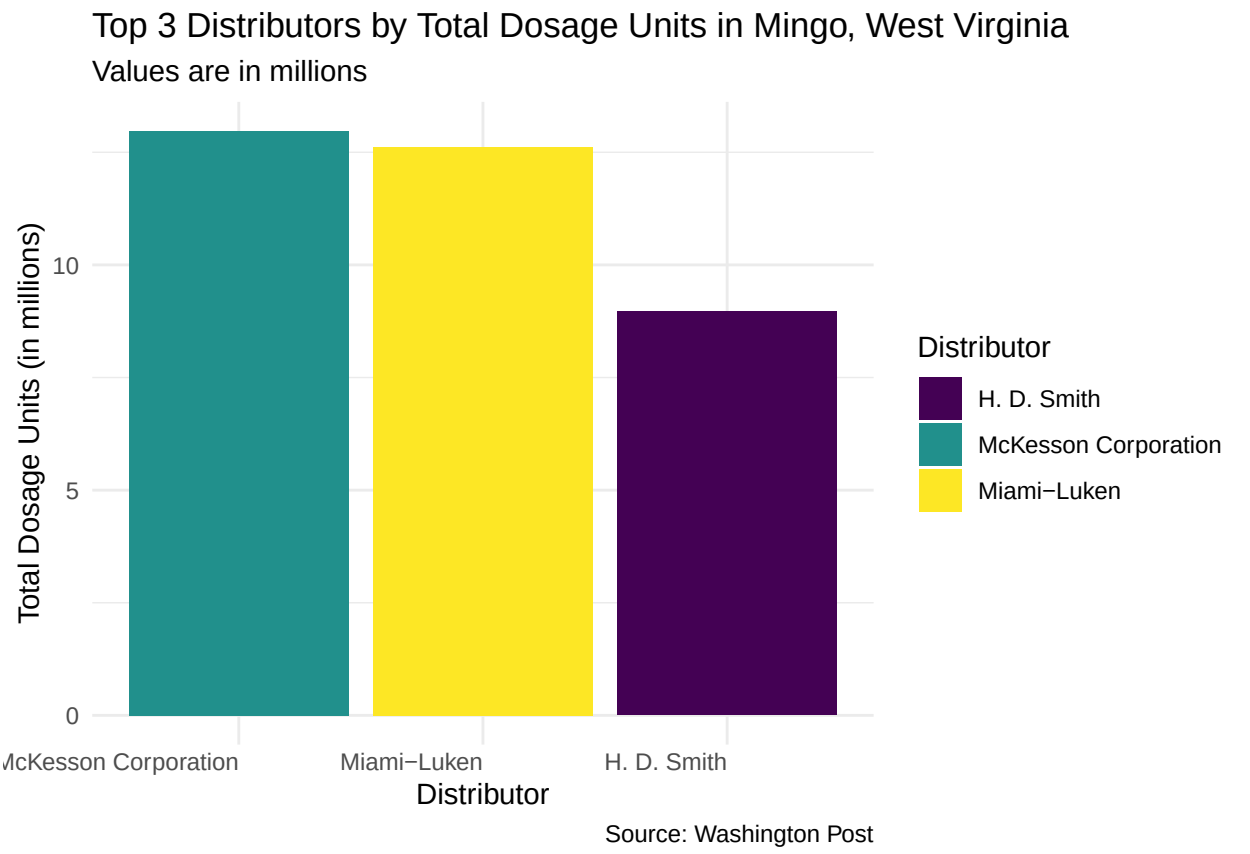
By formatting the data values for readability and presenting the information in a straightforward bar graph, the visualization ensures that the audience can easily understand the scale of distribution. This approach aligns with the ethical visualization principles of clarity and transparency. Additionally, the graph cites the Washington Post as the source of the data, which not only adds credibility but also allows viewers to consider the context and potential biases that come with the source. This acknowledgment of the data provenance is a critical aspect of responsible data presentation, as it maintains transparency and allows the audience to assess the information critically.

The data set was cleaned by formatting the values to make it readable to the audience to foster a clear understanding of the presented information. Furthermore, an essential aspect of ethical visualization lies in the transparent communication of data sources which has been sited below as from the Washington Post.

The insights that can be drawn from this graph are significant for public awareness. They highlight the involvement of these distributors in the supply chain of pharmaceuticals within a specific region, which could be a point of discussion in the context of public health, regulatory attention, or community impact. The ethical presentation of this data does not imply blame but rather provides a factual basis for understanding the distribution network's scale in Mingo County.

```
# Ethical Visualization
top3_ethical_example <- ggplot(data = head(distributor
[order(-distributor$total_dosage_unit), ], 3),
aes(x = reorder(Reporter_family, -total_dosage_unit),
y = total_dosage_unit,
fill = Reporter_family)) +
  geom_bar(stat = "identity") +
  scale_fill_viridis_d() +
  scale_y_continuous(labels =
  scales::comma_format(scale = 1e-6),
  breaks = scales::pretty_breaks(n = 5)) +
  labs(x = "Distributor",
       y = "Total Dosage Units (in millions)",  # y-axis label
       fill = "Distributor",
       title = "Top 3 Distributors by Total Dosage Units in Mingo, West Virginia",
       subtitle = "Values are in millions",
       caption = "Source: Washington Post") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 00, hjust = 1,
                  vjust = 1, margin = margin(t = 0.2, r = 0.2,
                  b = 0.2, l = 0.2)))

top3_ethical_example
```




```
ggsave("Top 3 Distributors.pdf", width = 11, height = 5.5)
```

Unethical Visualisation - Red team

The visualization created by the Red Team is a bar chart designed to showcase the comparison between different distributors based on total dosage units. At first glance, the chart seems to offer a straightforward comparison, but a closer look reveals several elements that might lead to misinterpretation or misleading conclusions.

This chart employs a cluttered design with bars positioned in a dodged layout, making it challenging to directly compare the distributors' performance accurately. The choice of color fill for each distributor further complicates the visual, potentially distracting from the actual data being represented. Although the labels for the distributor and total dosage units are clearly marked, the visual complexity introduced by these design choices does not facilitate an easy or ethical interpretation of the data.

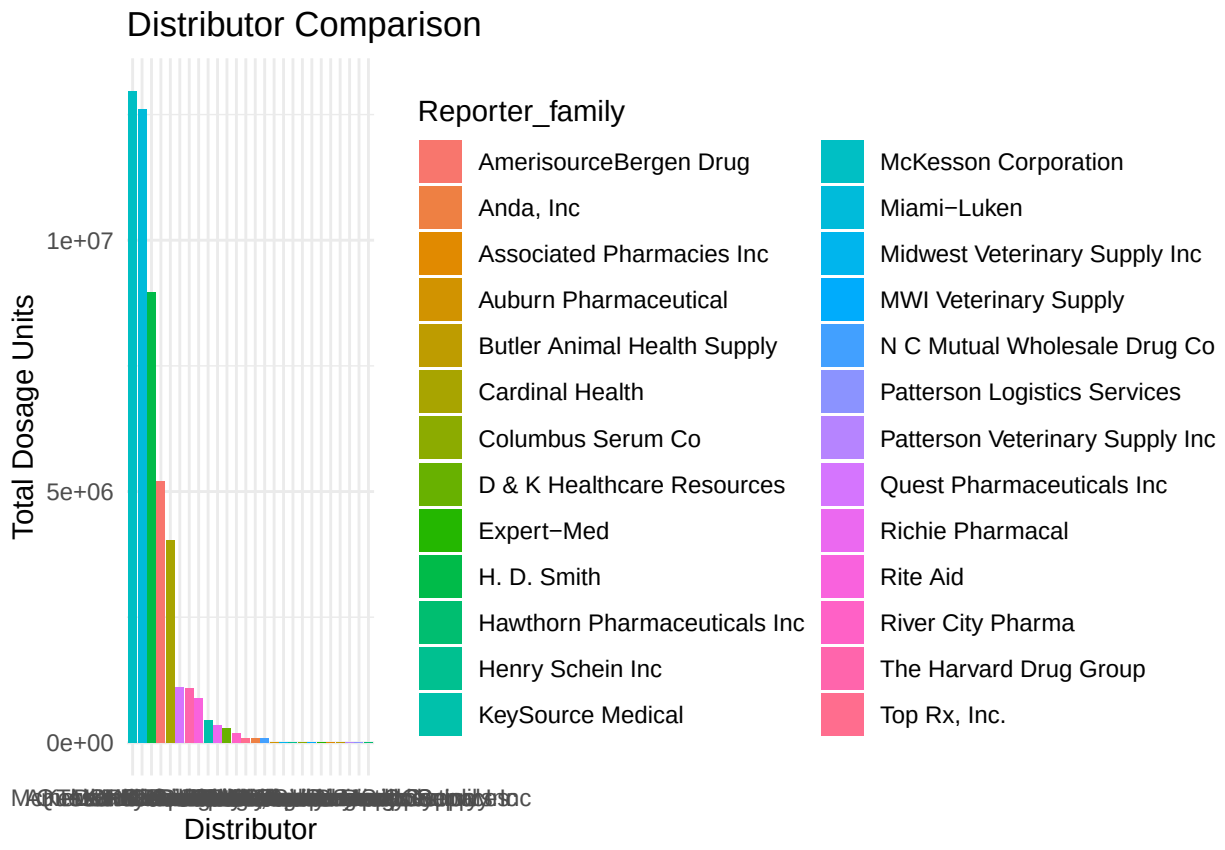
The chart claims to represent the total dosage units by each distributor, yet the presentation style chosen can obscure the data's true narrative. The use of a minimal theme does little to mitigate the visual clutter, and the overall effect may be an unethical representation of the data, where viewers could easily draw incorrect conclusions about the comparative performance of each distributor.

In crafting this visualization, the Red Team prioritize aesthetic complexity over clarity and accuracy, leading to a potentially unethical portrayal of important data. It serves as a reminder of the responsibility that comes with data visualization: to present information in a manner that is both accessible and truthful to the audience.

```r
# Unethical Visualization
library(ggplot2)

# Create a cluttered and misleading bar chart
unethical_plot <- ggplot(distributor,
aes(x = reorder(Reporter_family, -total_dosage_unit),
    y = total_dosage_unit, fill = Reporter_family)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Distributor Comparison",
       x = "Distributor",
       y = "Total Dosage Units") +
  theme_minimal()

# Display the plot in R Markdown
unethical_plot
```




```r
ggsave("unethical.pdf", width = 11, height = 5.5)
```