Here's my re-creation of the chart in R, complete with colours and styling to match the original created in SAS VA (note that my data – sourced from the New York Times – is quite different to that in the original chart for reasons unknown to me, but that doesn't seem to be important for illustrative purposes):

As a visualisation at least three things are wrong:

1. dodged bar charts are rarely effective for making comparisons over time – it's difficult for the eye to follow;
2. within each day's clump of bars, the counties are in a different order (highest to lowest, within the clump), reducing the meaning in the pattern in each clump;
3. the daily clumps of bars are not in chronological order.

Side issue – what is the correct word for what I am calling a "clump" of bars?

In my judgement, I've listed the worst mistake first; the other two seem to me secondary. Once you've decided not to use a line chart for this dataset, you've immediately lost the most powerful leverage a visulaisation can give you. This bar chart is never going to be much better than a simple table of numbers, no matter what sequence the bars are in.

Most critiques have focused on the third of the faults listed above, and assumed malicious intent to make it look like cases are going down faster than they are. However, I can easily see this error being made without thinking.

More often than not, the best thing to do visually with the categories in a bar or column chart *is* to order them by the values, so the reader can scan down or across the axis labels and get some information without even looking across at the bars. It's certainly much better than leaving the categories in (for example) alphabetical order. The convention that time series data goes from left to right should trump this rule about reordering your bars, but it's easy to imagine making a misjudgement there.

The second of the errors I found the most surprising – the counties appearing in a different order in each daily clump of bars. Like a few others I thought "how could you even do that?" But in my chart above I found it was actually quite easy, thanks to tools developed by the `tidytext` community to better present words in facets (for example, words associated with different topics as a result of topic modelling).

> Reordering your bars within each clump in a bar chart is easy with tidytext::reorder_within()

Here's the code to download county-level data from the New York Times and draw my re-creation of this bad chart. The trick is to have one variable `county` ordered by number of new cases *overall*, and use this for the fill and stroke colours; and a second variable `county2` that is reordered by cases *within* `date`. Of course, `date` itself has to be converted into a character string and then a factor for its own reordering ("mistake 3" in my original enumeration).

```
library(tidyverse)
library(scales)
library(tidytext)

#-------------prepare data-----------------

counties <- read_csv("https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv")

# find the top 5 counties
georgia5 <- counties %>%
  filter(state == "Georgia" & county != "Unknown") %>%
  group_by(county) %>%
  summarise(cases = max(cases)) %>%
  arrange(desc(cases)) %>%
  slice(1:5) %>%
```

```
  pull(county)

# dataset, all dates, for use later
d <- counties %>%
  filter(county %in% georgia5 & state == "Georgia") %>%
  group_by(county) %>%
  mutate(new_cases = cases - lag(cases)) %>%
  ungroup()
#-----------------theme, colours, titles---------------

# colours identified by putting the original image in Gimp
# and using the Color Picker:
fill_palette <- c(
  "Cobb" = "#5954b0",
  "DeKalb" = "#238a8f",
  "Fulton" = "#98863c",
  "Gwinnett" = "#965b31",
  "Hall" = "#2460ac"
)

stroke_palette <- c(
  "Cobb" = "#827cf8",
  "DeKalb" = "#42d1e1",
  "Fulton" = "#e5cd63",
  "Gwinnett" = "#e28f4d",
  "Hall" = "#3da2f2"
)

theme_set(  theme_minimal() +
              theme(panel.background = element_rect(fill = "#0f3051",
colour = NA),
                    panel.grid = element_blank(),
                    plot.background = element_rect(fill = "#0f3051"),
                    text = element_text(colour = "grey70"),
                    axis.text = element_text(colour = "grey70"),
                    axis.line = element_line(colour = "grey90"),
                    strip.text = element_text(colour = "grey80"),
                    legend.position = "top")
)

the_caption <- "Source: analysis by http://freerangestats.info with county-
level COVID-19 case data from New York Times"
title <- "Top 5 Counties in Georgia with the Greatest Number of
Confirmed COVID-19 Cases"
st <- str_wrap("Note that this chart is to illustrate poor visual
design choices and does not include
the most current data. It uses different data from the original
               from the Georgia Department of Public Health.", 120)

#--------draw bar chart---------------
d %>%
  filter(date >= as.Date("2020-04-27")) %>%
  filter(date <= as.Date("2020-05-09")) %>%
  mutate(date = fct_reorder(as.character(format(date, "%d%b%Y")),
-new_cases, .fun = sum),
         # reorder county for use for colour and legend:
         county = fct_reorder(county, -new_cases),
```

```
          # a new "county2" is reordered within date
          county2 = tidytext::reorder_within(county, -new_cases, within
= date)) %>%
  ggplot(aes(x = date, weight = new_cases, fill = county, colour =
county)) +
  geom_bar(position = "dodge", aes(group = county2)) +
  scale_colour_manual(values = stroke_palette) +
  scale_fill_manual(values = fill_palette) +
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) +
  labs(title = title,
       subtitle = st,
       colour ="",
       fill = "",
       x = "",
       y = "New cases per day",
       caption = the_caption)
```

Note the trick here is to force `geom_bar()` to use the `county2` variable for its grouping, while still leaving `county` to determine the stroke and fill colours. So, I'd say it is *easy* to do this with R. But I don't think you'd be likely to do it by mistake.

However, I'm still on the side of the original being a mistake rather than malign in intent. I think someone clicked the wrong button in SAS VA. Going up a level, I'm certain that the original sin in this case – the decision to use a bar chart rather than a line chart – was just a poor choice, not a conspiracy against the public.

A better visualisation

I couldn't leave this dataset without trying to visualise it properly. It turns out a single line chart is pretty messy and difficult to read. However, small multiples work nicely. With judicious use of layers, we can even have both the original data (as points) and a meaningful smoothing line (but not a projection forwards!) to help the eye:

With the appropriate chart for this sort of data, we can also show a meaningfully longer period of time too – a real problem with the bar chart.

Here's the R code to produce that chart, drawing on the same dataset (but without filtering it to just a few weeks of data) prepared earlier.

```
d %>%
  ggplot(aes(x = date, y = new_cases, colour = county)) +
  geom_point(alpha = 0.5) +
  geom_smooth(size = 1.5, se = FALSE, span = 0.5, method = "loess") +
  scale_colour_manual(values = stroke_palette) +
  facet_wrap(~county) +
  theme(legend.position = "none",
        panel.grid.major = element_line(colour = "#0f3081")) +
  labs(title = title,
       subtitle = "Improved visual presentation showing original data
and trend, and using full period of data available.",
       x = "",
       y = "New cases per day",
       caption = the_caption)
```