

Visualizing Annual Changes in Program Reach Across Several Countries

Introduction

On my team, one of the (rough) measures we use to assess the impact of a program is how many people that program has reached. This is interesting data that is rarely scrutinized, but recently I was asked to visualize which country programs showed an increase in TR from 2019 to 2020. As the scale of reach varies quite a bit across many countries, I found this to be an interesting request that could be standardized and used annually for a quick glance at each country's performance.

My first thought was to use a slope graph, which is often used to visualize comparisons. These can get a bit messy with too many variables, but faceting them is one option to clean things up a bit. Then a friend suggested using a bar graph that visualizes the net difference of program reach between 2020 and 2019. This is also a really great option, though it can get a bit cluttered with too many variables. Below I'll walk you through the process of creating both so that you can decide which is more useful for your own project. First let's load in our package and then we will generate sample data to visualize. We'll just need ggplot and dplyr from the tidyverse package for this one.

Generate Sample Data

For this exercise, we need the country variable, 2019 and 2020 data. For simplicity purposes, I decided to generate sample data rather than use existing data. You can use `sprintf` to generate a numbered variable, which in this case will create sample countries numbered from 1 to 8. Then, use `sample` to generate 8 random numbers for 2019 and 8 random numbers for 2020. Pipe all this into a data frame using `data.frame`.

```
set.seed(3232021)

country <- c(sprintf("country%01d", seq(1,8)))

tr_2019 <- sample(0:10000, 8, replace=TRUE)

tr_2020 <- sample(0:10000, 8, replace=TRUE)

tr <- data.frame(country, tr_2019, tr_2020)
```

Transform Data

There are three main steps to prepare the data for visualizations. The first is an addition of a column that will contain the net difference between 2020 and 2019. We'll use this for the second visualization. The second is the addition of a column that will represent whether or not there was an increase in program reach from 2019 to 2020, which will be used as a fill to highlight increases or decreases. You can use `mutate` to create a new column and `ifelse` to essentially say, "If the 2020 column is greater than the 2019 column, generate an 'increase' observation, otherwise generate a 'decrease'". The third step is to use `gather` to transform the data by creating 'year' and 'reach' variables from the columns with the annual reach observations

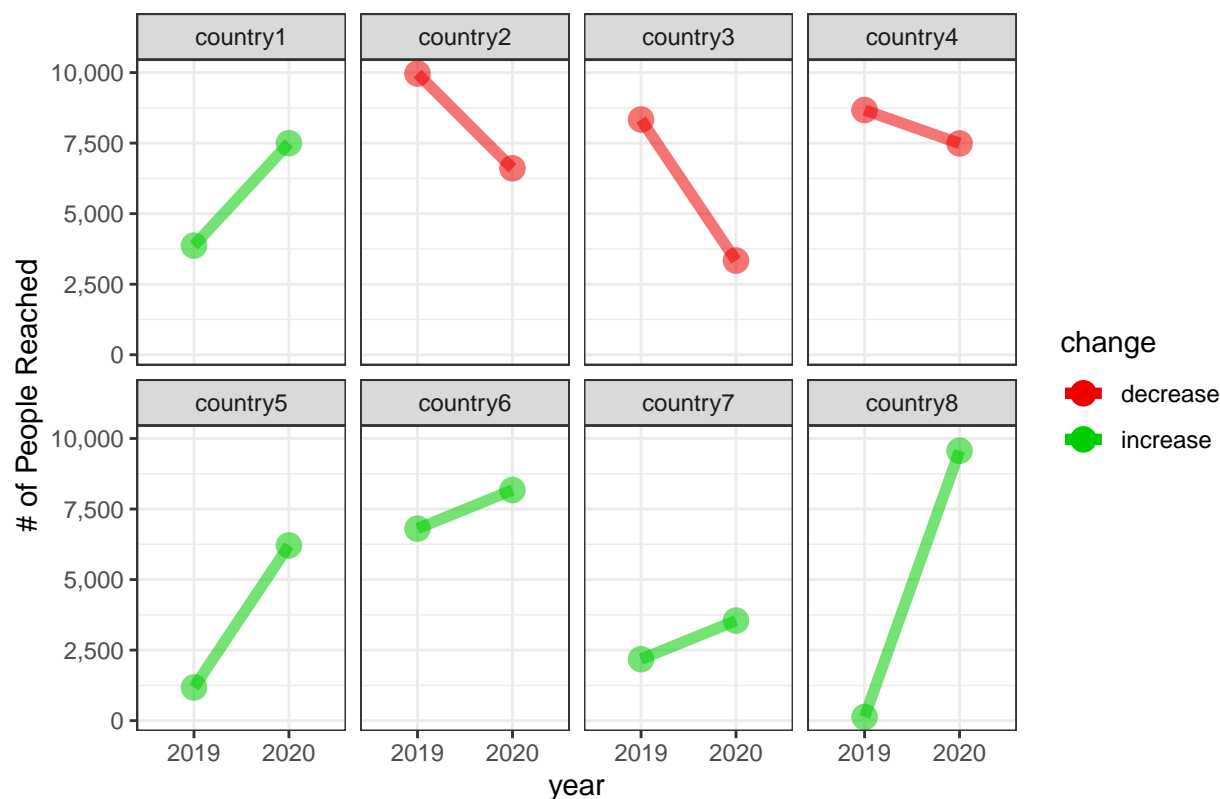
```
tr_trans <- tr %>%
  mutate(net_diff = c(tr$tr_2020 - tr$tr_2019)) %>%
  mutate(change = ifelse(tr$tr_2020 > tr$tr_2019, "increase", "decrease")) %>%
  rename('2020' = tr_2020, '2019' = tr_2019) %>%
  gather(year, reach, 2:3)
```

Visualize Data

The first visualization is a faceted slope graph. I originally tried to visualize this on one graph, but because my data had 22 countries with vastly different scales, it came out looking a bit messy. So, I decided to facet the data by country. If you have just a few observations, you may want to consider putting all observations on one graph, though. I think it is also important to generate bold colors to represent the increases and decreases.

```
ggplot(tr_trans, aes(x = year, y = reach, group = country)) +
  geom_line(aes(color = change, alpha = 1), size = 2) +
  geom_point(aes(color = change, alpha = 1), size = 4) +
  scale_alpha(guide = 'none') +
  facet_wrap(. ~ country, ncol = 4) +
  scale_color_manual(values=c("red2", "green3")) +
  theme(axis.text.x = element_text(size = 10, angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::comma) +
  ggtitle("Five of Eight Countries Saw an Increase in Program Reach in 2020") +
  theme(plot.title = element_text(hjust = 0.5, margin=margin(10,0,15,0))) +
  ylab("# of People Reached") +
  theme(axis.title.y = element_text(margin = margin(t = 0, r = 20, b = 0, l = 0))) +
  theme(axis.title.x = element_text(margin = margin(t = 10, r = 0, b = 10, l = 0))) +
  theme(panel.border = element_blank()) +
  theme_bw()
```

Five of Eight Countries Saw an Increase in Program Reach in 2020



The second visualization is a bar graph that uses color and `geom_hline` to highlight net differences between 2020 and 2019. While this option doesn't show the variation in program reach scales, you'll be able to see the overall increase or decrease between the two years. As there is no true order to the countries, I decided to use `reorder` to order the bars by net difference instead of by country.

```
ggplot(tr_trans, aes(reorder(country, -net_diff), net_diff, group = change)) +
  geom_col(stat = "identity", aes(fill = change, alpha = 1)) +
  scale_alpha(guide = 'none') +
  scale_fill_manual(values=c("red2", "green3")) +
  theme(axis.text.x = element_text(size = 10, angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::comma) +
  ggtitle("Five of Eight Countries Saw an Increase in Program Reach in 2020") +
  theme(plot.title = element_text(hjust = 0.5, margin=margin(10,0,15,0))) +
  ylab("Net Difference Between 2020 and 2019") +
  xlab("") +
  theme(axis.title.y = element_text(margin = margin(t = 0, r = 20, b = 0, l = 0))) +
  theme(axis.title.x = element_text(margin = margin(t = 10, r = 0, b = 10, l = 0))) +
  theme(panel.border = element_blank()) +
  geom_hline(yintercept=0) +
  theme_bw() +
  coord_flip()
```

Five of Eight Countries Saw an Increase in Program Reach in 2020

