US-murders ggplot

Prasham Bhuta

08/04/2020

Generating a plot that describes the US murder rate for all states based on the region.

The dataset is a part of the learning at Harvardx's Data Science Course.

Loading required packages

```
library(dslabs)
library(tidyverse)
library(ggthemes)
library(ggrepel)
```

Create a plot object with required data.

```
p <- murders %>% ggplot(aes(population / 10^6, total, label=abb), height = 6, width = 16)
```

Adding text geometry

```
p <- p + geom_text_repel(nudge_x = 0.04, nudge_y = 0.04)
```

Changing axis scale to be logarithmic

```
p <- p + scale_x_log10() + scale_y_log10()
```

Adding labels & titles

```
p <- p + xlab("Population in millions (log)") +
     ylab("Total murders (log)") +
     ggtitle("US murders across states filtered by region")
```

Calculate the average murder rate

This value will be plotted as line across the plot

```
avg_rate <- murders %>% summarise(rate = sum(total) / (sum(population)) * 10^6) %>% .$rate
```

Adding abline for the `avg_rate`

```
p <- p + geom_abline(intercept = log10(avg_rate), color="orange", lty =2, size =1)
```

Adding points as state with region as filter

```
p <- p + geom_point(aes(col = region), size = 3)
```

Renaming the label 'region'

```
p <- p + scale_color_discrete(name = "Regions in US")
```

Adding theme `economist` to our plot and improving font sizes

```
p <- p + theme_economist() + theme(text = element_text(size = 16))
```

The desired plot