Replicating Plots: Oil Pathways

Stomperusa

2020-02-08

### Intro

I was inspired by this graphic on Yahoo Finance. In the TidyTuesday spirit, I set out to recreate it using ggplot2. There were a few things I would have to figure out - how to apply ggtext to an annotation, how to draw a curved arrow, and how to create a pinhole like border around the graphic. That last point, well, I haven’t figured that one out yet and so I am looking for ideas from the R community. The graphic requires very little data that is easy enough to type out from scratch, but I found the source data in a pdf file and couldn’t resist using tabulizer to extract it.



library(tabulizer)  
library(tidyverse)  
library(sf)  
library(rnaturalearth)  
library(rnaturalearthdata)  
library(ggtext)  
library(here)

### Using Tabulizer to extract the source data

If you are looking to extract the data directly from the source pdf file, here is the code for that.

# Not Run  
 # This is the source location of the data in a pdf file  
#location <- str\_c("https://www.eia.gov/beta/international/analysis\_includes/",  
# "special\_topics/World\_Oil\_Transit\_Chokepoints/wotc.pdf")  
  
 # Use tabulizer to extract the relevant portion of the data table on page 2 of the pdf  
#extract <- extract\_areas(file =location, pages = 2, resolution = 100)  
  
 # Set up the data as a data frame and then isolate just year 2016 for the graphic  
#df <- as\_tibble(extract[[1]][2:nrow(extract[[1]]),])  
#names(df) <- extract[[1]][1,]  
  
#data <- df %>%   
# select("Location", Barrels = "2016") %>%   
# mutate(Barrels = as.numeric(Barrels))  
  
#coord <- tribble(  
#~Location, ~Lat, ~Lon,  
#"Strait of Hormuz", 26.5667, 56.2500,  
#"Strait of Malacca", 1.4300, 102.8900,  
#"Suez Canal and SUMED Pipeline", 30.4550, 32.3500,  
#"Bab el-Mandeb", 12.5833, 43.3333,  
#"Danish Straits", 56.0000, 11.0000,  
#"Turkish Straits", 40.7225, 28.2247,  
#"Panama Canal", 9.38743, -79.91863,  
#"Cape of Good Hope", -34.3548, 18.4698  
#)  
  
#data <- data %>%   
# left\_join(coord, by = "Location")  
 #The source data has "Turkish Straits" but Yahoo Finance labeled it "Bosporus".  
#data[data$Location =="Turkish Straits",1] <- "Bosporus"  
  
#write\_csv(data, "oil\_pathway\_data.csv")

Or you can get the prepped data set directly from my github at: <https://github.com/stomperusa/VisualReproduction/tree/master/OilPathways>

data <- read\_csv("oil\_pathway\_data.csv")

### The Graph

I have just recently discovered ggtext and the ability it introduces to control formats within a label. I was expecting to set the title of the graph as an annotation given that it sits within the plot. My first learning point was that ggtext has geom\_richtext() for this purpose.

title <- "<b style='color:#FF0000'>IRAN</b> <b style='color:#FFFFFF'>BORDERS<br> KEY OIL PATHWAY</b>"

I used rnaturalearth for the map. The original may be based on another map projection given the orientation differences apparent at the top of the graphic, around Greenland for example.

world <- ne\_countries(scale = "medium", returnclass = "sf")

The geom\_curve was a new discovery for me as well. How to get the pinhole effect and the shading around the borders eludes me. Reach out if you know how to resolve.

Here is my stab at reproducing the graphic. Obviously there is a bit more fine tuning I could do to get it closer to the original, but I got what I wanted out of the exercise.

plot <- ggplot(data = world) +  
 geom\_sf(fill = "blue", color = "black", size = 0.2) +  
 geom\_sf(data = subset(world, geounit == "Iran"), fill = "red", color = "black", size = 0.2) +  
 geom\_point(data = data, aes(x =Lon, y=Lat, size = Barrels),   
 shape = 21, fill = "goldenrod2", color = "white", show.legend = F) +  
 geom\_text(data = data, aes(x =Lon, y=Lat, label = Location), color = "white", size = 3.5,  
 vjust = -1, hjust = c(-0.3, 1.25, 1.1, 1.2, -0.15, -0.15, -0.15, 1.2),   
 fontface = "bold") +  
 geom\_text(data = data, aes(x =Lon, y=Lat, label = Barrels), color = "goldenrod2", size = 3.5,  
 vjust = .5, hjust = c(-1.0, 2.7, 2.0, 2.0, -1.0, -1.0, -0.9, 2.1),  
 fontface = "bold") +  
 scale\_size\_continuous(range = c(2, 15)) +  
 annotate("text", x = -60, y = 60, label = "MILLION BARRELS\nOF OIL MOVE PER\nDAY, 2016 DATA", size = 4.0,  
 color = "goldenrod2", fontface= "bold") +  
 annotate("text", x = -60, y = -60, label = "SOURCE U.S. ENERGY INFORMATION ADMINISTRATION", size = 2.3,  
 color = "white")+  
 geom\_richtext(x = 20, y = 80, label = title, size = 9.0, fill = NA, label.color = NA) +  
 geom\_curve(aes(x = 73, y = 80, xend = 65, yend = 38), color = "red", arrow = arrow(type = "open",  
 length = unit(0.15, "inches")), curvature = -0.75, angle = 100, ncp =10) +   
 coord\_sf(xlim = c(-100.00, 120.00), ylim = c(-65, 95), expand = F) +  
 theme(panel.background = element\_rect(fill = "black"),  
 panel.grid = element\_blank(), axis.title = element\_blank(), axis.ticks = element\_blank(),  
 axis.text = element\_blank(), aspect.ratio = 1)   
  
#ggsave(here::here("oil\_pathways.png"), plot, dpi = 320, width = 20, height = 20, units = "cm")

Here’s the link to the article that contains the original:

<https://finance.yahoo.com/news/what-us-iran-tensions-could-mean-for-oil-prices-193905568.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAADp4Vt8w3NAGpqbj8B0Lj3oLzHjF-R9YqWXZKXiKiabBmEv_8GNYKJeLGytcLIHL1fLnn9Rv886i_Vr07iqAUu4fuZrpm-5goU7ulysC0QjUSEUxWfs1Ax_zinvN3DezuR3zAsWmwSClEXkSGXu9_d-wmCe22K4GKnRjQq7_1d3p>