Improving Graphics

Nabina Khadka

2023-03-15

## Libraries

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(plotly)

##   
## Attaching package: 'plotly'  
##   
## The following object is masked from 'package:ggplot2':  
##   
## last\_plot  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following object is masked from 'package:graphics':  
##   
## layout

library(scatterplot3d)

## Loading Files

load("TFR\_long.Rdata")  
load("IMR\_long.RData")

## Checking for MisMatches

miss = anti\_join(IMR\_long,  
 TFR\_long,  
 by = c("country\_name","year"))  
nrow(miss)

## [1] 0

## Joining two data sets

IMR\_TFR = IMR\_long %>%   
 left\_join(TFR\_long,  
 by = c("country\_name",  
 "country\_code",  
 "year"))  
glimpse(IMR\_TFR)

## Rows: 16,226  
## Columns: 5  
## $ country\_name <chr> "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Ar…  
## $ country\_code <chr> "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "…  
## $ year <dbl> 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 196…  
## $ IMR <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…  
## $ TFR <dbl> 4.820, 4.655, 4.471, 4.271, 4.059, 3.842, 3.625, 3.417, 3…

## Scatter Plot

g = IMR\_TFR %>%   
 ggplot(aes(IMR,TFR,group = country\_name)) +  
 geom\_point(size = .1)  
ggplotly(g)

## Improving the plot

The scatter plot shows that there is a chunk of data at the start and it is difficult to explain and examize the data. To improve on the graph, let us apply few more data visualization techniques.

Let us start.

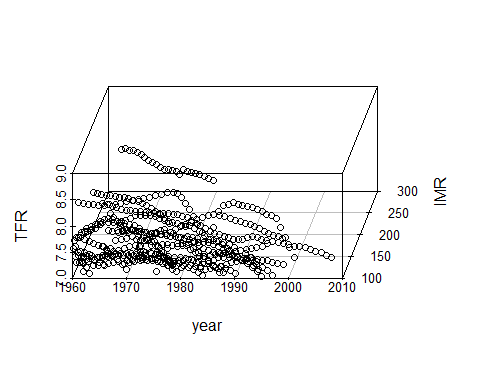
## ScatterPlot 3D

data <- IMR\_TFR %>%  
 filter(IMR > 100.0000 ,   
 TFR > 7.0000 )  
  
glimpse(data)

## Rows: 546  
## Columns: 5  
## $ country\_name <chr> "Afghanistan", "Afghanistan", "Afghanistan", "Afghanistan…  
## $ country\_code <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "…  
## $ year <dbl> 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 197…  
## $ IMR <dbl> 237.5, 233.2, 229.2, 225.3, 221.5, 217.7, 213.8, 209.9, 2…  
## $ TFR <dbl> 7.284, 7.292, 7.302, 7.304, 7.305, 7.320, 7.339, 7.363, 7…

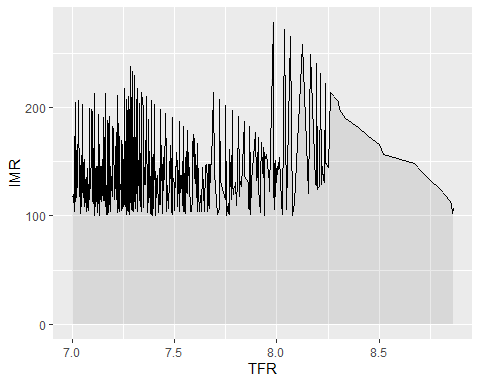
scatterplot3d(data[,3:5] , angle = 75)

## Warning: Unknown or uninitialised column: `color`.



## Area Graph

ggplot(data , aes(x = TFR , y = IMR)) +  
 geom\_area(alpha = 0.1) +  
 geom\_line()



## Density Plot

ggplot(data , aes(x= TFR , y = IMR )) +  
 geom\_bin2d() +  
 theme\_bw()

