Graph I - SA

Sachin Shubham

3/6/2021

library(ggplot2)  
library(MASS)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble 3.0.5 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1  
## v purrr 0.3.4

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::select() masks MASS::select()

library(dplyr)  
library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

library(ggtext)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(gtable)  
library(cowplot)  
library(grid)  
library(egg)  
library(e1071)  
setwd('D:/Statistical Methods/Project')  
dataFromCSV<-read.csv("ArticleLevel-RegData-ALLSA\_Xc\_1\_NData\_655386\_LONGXCIP2.csv")  
  
fd<-dataFromCSV %>%  
 group\_by(Yp, NSAp, Zp>0) %>%  
 summarise(  
 cnt = n()  
 ) %>%  
 filter(  
 Yp>1979  
 )

## `summarise()` has grouped output by 'Yp', 'NSAp'. You can override using the `.groups` argument.

yrcnt = NULL  
years = 1980:2018  
  
for (i in 1:length(years)){  
 hht <- filter(fd, Yp==i+1979)  
 yrcnt[i] <- ( sum(hht[4]))  
}  
  
  
  
############################################  
# FIGURE 2A - 1st Graph SA Categories  
############################################  
  
# Specify years and categories to analyze  
years <- 1980:2018  
cats <- 1:4  
  
  
# Create matrix to store SA papers in specified categories by years  
yrSAcat <- matrix(data=NA,nrow= length(cats)\*2, ncol=length(years) )  
  
# Create vectors to store total SA papers with below average citations in specified categories by years  
yrSAtotNEG <- matrix(data=NA, nrow = 1, ncol = length(years))  
  
# Create vectors to store total SA papers with above average citations in specified categories by years  
yrSAtotPOS <- matrix(data=NA, nrow = 1, ncol = length(years))  
  
  
  
# create negative & positive arrays to store values from yrSAcat  
negarr\_SA <- (cats-1)\*2+1  
posarr\_SA <- (cats)\*2  
  
  
  
  
  
for (i in 1:length(years)){  
 for (j in 1:(length(cats)\*2)){  
 rr <- filter(fd, Yp == (i+1979) )  
 yrSAcat[j,i] <- rr[[j,4]]  
 }  
 yrSAtotNEG[1,i] <- sum(yrSAcat[negarr\_SA,i])  
 yrSAtotPOS[1,i] <- sum(yrSAcat[posarr\_SA,i])  
   
}  
  
  
  
colnames(yrSAcat) <- c(1980:2018)  
rownames(yrSAcat) <- c("SA1 Zp-", "SA1 Zp+", "SA2 Zp-", "SA2 Zp+", "SA3 Zp-", "SA3 Zp+", "SA4 Zp-", "SA4 Zp+")  
colnames(yrSAtotNEG) <- paste(1980:2018, "Neg Cite.")  
colnames(yrSAtotPOS) <- paste(1980:2018, "Pos Cite.")  
  
  
  
SApercTotNEG <- matrix(data=NA, nrow = length(cats), ncol = length(years))  
SApercTotPOS <- matrix(data=NA, nrow = length(cats), ncol = length(years))  
  
  
  
for (i in 1:length(years)){  
 for (j in 1:length(cats) ) {  
 SApercTotNEG[j,i] <- yrSAcat[((j-1)\*2+1),i]/yrSAtotNEG[1,i]  
 SApercTotPOS[j,i] <- yrSAcat[(j\*2),i]/yrSAtotPOS[1,i]  
 }  
}  
  
  
  
#colnames(SApercTot) <- c(1980:2018)  
rownames(SApercTotNEG) <- c("SA1 Zp- ", "SA2 Zp- ", "SA3 Zp- ", "SA4 Zp- ")  
rownames(SApercTotPOS) <- c("SA1 Zp+ ", "SA2 Zp+ ", "SA3 Zp+ ", "SA4 Zp+ ")  
  
  
  
  
years <- 1980:2018  
  
#PLOTTING for SA  
  
p1 = data.frame(years=years, Fraction = (SApercTotPOS[1,]))  
p2 = data.frame(years=years, Fraction = (SApercTotPOS[2,]))  
p3 = data.frame(years=years, Fraction = SApercTotPOS[3,])  
p4 = data.frame(years=years, Fraction = (SApercTotPOS[4,]))  
  
  
n1 = data.frame(years=years, Fraction = (SApercTotNEG[1,]))  
n2 = data.frame(years=years, Fraction = (SApercTotNEG[2,]))  
n3 = data.frame(years=years, Fraction = SApercTotNEG[3,])  
n4 = data.frame(years=years, Fraction = (SApercTotNEG[4,]))  
  
e1<-ggplot()+geom\_line(data=p1, mapping = aes(x=years,y=Fraction), size=0.5, color="grey30") + geom\_line(data=p2, mapping = aes(x=years,y=Fraction), size=0.5, color="grey69") +  
 geom\_line(data=p3, mapping = aes(x=years,y=Fraction), size=0.5, color="royalblue") + geom\_line(data=p4, mapping = aes(x=years,y=Fraction), size=0.5, color="blue") +  
 geom\_line(data=n1, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey30") + geom\_line(data=n2, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey69") +  
 geom\_line(data=n3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="royalblue") + geom\_line(data=n4, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="blue")  
   
e2<-e1+theme\_bw() + theme(panel.border = element\_rect(colour = "black", fill=NA, size=1), panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"),  
 axis.text.x = element\_text(size=10))   
e3<-e2+scale\_x\_continuous(limits=c(1980, 2020), breaks=seq(1980, 2020, 5))+scale\_y\_continuous(limits=c(0.0, 0.6), breaks=seq(0.0, 0.6, 0.2))  
  
  
e4<-e3+labs(title = "", x = "", y = "")  
  
  
e5<-e4+ scale\_color\_manual(name = "Y series")  
  
e6<-e5+theme(legend.title = element\_text(size=12, color = "salmon",   
 face="bold"),  
 legend.justification=c(0,1),   
 legend.position=c(0.05, 0.95),  
 legend.background = element\_blank(),  
 legend.key = element\_blank()) +   
 labs(subtitle="Cross-Topic(SA)")  
  
  
e6

