Untitled

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 - 2nd Graph CIP Categories  
######################################  
  
# Specify years and categories to analyze  
years <- 1980:2018  
cats <- 1:4  
  
# Store CIP values of all categories from CSV data  
CIPp<-dataFromCSV %>%  
 group\_by(Yp, NCIPp, Zp>0) %>%  
 summarise(  
 cnt = n()  
 ) %>%  
 filter(  
 Yp>1979  
 )

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

# Create matrix to store CIP papers in specified categories by years  
yrCIPcat <- matrix(data=NA,nrow= length(cats)\*2, ncol=length(years) )  
  
# Create vectors to store total CIP papers with below average citations in specified categories by years  
yrCIPtotNEG <- matrix(data=NA, nrow = 1, ncol = length(years))  
# Create vectors to store total CIP papers with above average citations in specified categories by years  
yrCIPtotPOS <- matrix(data=NA, nrow = 1, ncol = length(years))  
  
# create negative & positive arrays to store values from yrCIPcat  
negarr\_CIP <- (cats-1)\*2+1  
posarr\_CIP <- (cats)\*2  
  
zer <- matrix(data = 0, nrow = 8, ncol = 4)  
tt2 <- tibble(Yp =(Yp = zer[1:8]), NCIPp=as.integer(0),   
 'Zp > 0'=FALSE, cnt= as.integer(0))  
  
for (i in 1:length(years)){  
 tt <- filter(CIPp, Yp == (i+1979) )  
 if (dim(tt)[1] < 8){  
 tt2 <- tibble(Yp = as.integer(zer[(8-dim(tt)[1]):8]), NCIPp=as.integer(0),   
 'Zp > 0'=FALSE, cnt= as.integer(0))  
 tt <- tt %>% full\_join(tt2,by=NULL)  
 }  
 for (j in 1:dim(tt)[1]){  
 if (j>8){  
 next  
 }  
   
 yrCIPcat[j,i] <- tt[[j,4]]  
   
 }   
 yrCIPtotNEG[1,i] <- sum(yrCIPcat[negarr\_CIP,i])  
 yrCIPtotPOS[1,i] <- sum(yrCIPcat[posarr\_CIP,i])  
}

## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")

## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")  
## Joining, by = c("Yp", "NCIPp", "Zp > 0", "cnt")

colnames(yrCIPcat) <- c(1980:2018)  
rownames(yrCIPcat) <- c("CIP1 Zp-", "CIP1 Zp+", "CIP2 Zp-", "CIP2 Zp+", "CIP3 Zp-", "CIP3 Zp+", "CIP4 Zp-", "CIP4 Zp+")  
colnames(yrCIPtotNEG) <- paste(1980:2018, "Neg Cite.")  
colnames(yrCIPtotPOS) <- paste(1980:2018, "Pos Cite.")  
  
CIPpercTotNEG <- matrix(data=NA, nrow = length(cats), ncol = length(years))  
CIPpercTotPOS <- matrix(data=NA, nrow = length(cats), ncol = length(years))  
  
for (i in 1:length(years)){  
 for (j in 1:length(cats) ) {  
 CIPpercTotNEG[j,i] <- yrCIPcat[((j-1)\*2+1),i]/yrCIPtotNEG[1,i]  
 CIPpercTotPOS[j,i] <- yrCIPcat[(j\*2),i]/yrCIPtotPOS[1,i]  
 }  
}  
  
colnames(CIPpercTotNEG) <- c(1980:2018)  
colnames(CIPpercTotPOS) <- c(1980:2018)  
  
rownames(CIPpercTotNEG) <- c("CIP1 Zp- %", "CIP2 Zp- %", "CIP3 Zp- %", "CIP4 Zp- %")  
rownames(CIPpercTotPOS) <- c("CIP1 Zp+ %", "CIP2 Zp+ %", "CIP3 Zp+ %", "CIP4 Zp+ %")  
  
years <- 1980:2018  
  
  
#Plotting for CIP  
  
  
  
  
cp1 = data.frame(years=years, Fraction = (CIPpercTotPOS[1,]))  
cp2 = data.frame(years=years, Fraction = (CIPpercTotPOS[2,]))  
cp3 = data.frame(years=years, Fraction = (CIPpercTotPOS[3,]))  
cp4 = data.frame(years=years, Fraction = (CIPpercTotPOS[4,]))  
  
  
cn1 = data.frame(years=years, Fraction = (CIPpercTotNEG[1,]))  
cn2 = data.frame(years=years, Fraction = (CIPpercTotNEG[2,]))  
cn3 = data.frame(years=years, Fraction = (CIPpercTotNEG[3,]))  
cn4 = data.frame(years=years, Fraction = (CIPpercTotNEG[4,]))  
  
ce1<-ggplot()+geom\_line(data=cp1, mapping = aes(x=years,y=Fraction), size=0.5, color="grey30") + geom\_line(data=cp2, mapping = aes(x=years,y=Fraction), size=0.5, color="grey69") +  
 geom\_line(data=cp3, mapping = aes(x=years,y=Fraction), size=0.5, color="royalblue") + geom\_line(data=cp4, mapping = aes(x=years,y=Fraction), size=0.5, color="blue") +  
 geom\_line(data=cn1, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey30") + geom\_line(data=cn2, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey69") +  
 geom\_line(data=cn3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="royalblue") + geom\_line(data=cn4, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="blue")  
  
ce2<-ce1+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))   
ce3<-ce2+scale\_x\_continuous(limits=c(1980, 2020), breaks=seq(1980, 2020, 5))+scale\_y\_continuous(limits=c(0.0, 1.0), breaks=seq(0.0, 1.0, 0.2))  
  
  
ce4<-ce3+labs(title = "", x = "", y = "")  
  
  
ce5<-ce4 +annotation\_custom(  
 ggplotGrob(ce4+annotation\_logticks() +  
 scale\_y\_log10(limits = c(NA,0.1))+scale\_x\_continuous(limits=c(1980, 2020), breaks=seq(1980, 2020, 10))),  
 xmin = 1983, xmax = 2007, ymin = 0.23, ymax = 0.8)

## Scale for 'y' is already present. Adding another scale for 'y', which will  
## replace the existing scale.

## Scale for 'x' is already present. Adding another scale for 'x', which will  
## replace the existing scale.

## Warning: Transformation introduced infinite values in continuous y-axis  
  
## Warning: Transformation introduced infinite values in continuous y-axis  
  
## Warning: Transformation introduced infinite values in continuous y-axis  
  
## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 39 row(s) containing missing values (geom\_path).

## Warning: Removed 24 row(s) containing missing values (geom\_path).

## Warning: Removed 39 row(s) containing missing values (geom\_path).

## Warning: Removed 13 row(s) containing missing values (geom\_path).

ce6<-ce5+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-Discipline(CIP)")  
  
ce6

