

Course Project Part 1

Sachin Shubham

3/8/2021

R Markdown

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

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

## v tibble  3.1.0      v dplyr    1.0.4
## v tidyrr   1.1.2      v stringr  1.4.0
## v readr    1.4.0      vforcats  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)

# Choose ArticleLevel-RegData-ALLSA_Xc_1_NData_655386_LONGXCIP2.csv
dataFromCSV<-read.csv(file.choose())

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]))
}

yrcnt

## [1] 1486 1745 1983 2306 2473 2982 3237 3839 4388 4659 4993 5681
## [13] 6195 6843 7398 8530 8888 10195 10974 11834 12780 14289 15955 18041
## [25] 20846 22941 25720 27416 29938 32313 33634 36421 36828 36985 36653 35973
## [37] 35150 33516 31371

#####
# 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+ ")

#####
# 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 year
yrCIPtotNEG <- matrix(data=NA, nrow = 1, ncol = length(years))
# Create vectors to store total CIP papers with above average citations in specified categories by year
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")
## 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-",
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+ %")

#####
# FIGURE 2A - 3rd Graph SA & CIP Categories
#####

# Specify years and categories to analyze
years <- 1980:2018
cats <- 1:4
```

```

SA_CIP1 <- dataFromCSV %>%
  group_by(Yp, NSAp, NCIPp, Zp>0) %>%
  summarise(
    cnt = n()
  ) %>%
  filter(NCIPp == 1 &
         NSAp == 1 & Yp>1979)

```

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

```

SA_CIP2 <- dataFromCSV %>%
  group_by(Yp, NSAp, NCIPp, Zp>0) %>%
  summarise(
    cnt = n()
  ) %>%
  filter(NCIPp == 2 &
         NSAp == 2 &
         Yp>1979)

```

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

```

SA_CIP3 <- dataFromCSV %>%
  group_by(Yp, NSAp, NCIPp, Zp>0) %>%
  summarise(
    cnt = n()
  ) %>%
  filter(NCIPp == 3 &
         NSAp == 3 &
         Yp>1979)

```

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

```

SA_CIP4 <- dataFromCSV %>%
  group_by(Yp, NSAp, NCIPp, Zp>0) %>%
  summarise(
    cnt = n()
  ) %>%
  filter(NCIPp == 4 &
         NSAp == 4 &
         Yp>1979)

```

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

```
SA_CIP <- SA_CIP1 %>% full_join(SA_CIP2, by = NULL) %>% full_join(SA_CIP3, by = NULL) %>% full_join(SA_CIP4, by = NULL)
```

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

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

```

# Create matrix to store SA_CIP papers in specified categories by years
yrSA_CIPcatNEG1 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=FALSE, cnt= 0)
yrSA_CIPcatPOS1 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=TRUE, cnt= 0)
yrSA_CIPcatNEG2 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=FALSE, cnt= 0)
yrSA_CIPcatPOS2 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=TRUE, cnt= 0)
yrSA_CIPcatNEG3 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=FALSE, cnt= 0)
yrSA_CIPcatPOS3 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=TRUE, cnt= 0)
yrSA_CIPcatNEG4 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=FALSE, cnt= 0)
yrSA_CIPcatPOS4 <- tibble(Yp = 0, NSAp=0, NCIPp=0, `Zp > 0`=TRUE, cnt= 0)

# Create vectors to store total CIP papers with below average citations in specified categories by year
yrSA_CIPtotNEG <- matrix(data=NA, nrow = 1, ncol = length(years))

# Create vectors to store total CIP papers with above average citations in specified categories by year
yrSA_CIPtotPOS <- matrix(data=NA, nrow = 1, ncol = length(years))

# tmp1 <- tibble(Yp = as.integer(Yp = zer[1:8]), NCIPp=as.integer(0), 'Zp > 0'=FALSE, cnt= as.integer(0))
for (i in 1:length(years) ){

  # Filter SA_CIP of all categories
  tmp1 <- filter(SA_CIP, Yp == (i+1979) )

  yrSA_CIPcatNEG1 <- yrSA_CIPcatNEG1 %>% full_join(filter(tmp1, (NCIPp == 1 | NSAp == 1) & `Zp > 0` == FALSE))
  yrSA_CIPcatPOS1 <- yrSA_CIPcatPOS1 %>% full_join(filter(tmp1, (NCIPp == 1 | NSAp == 1) & `Zp > 0` == TRUE))

  yrSA_CIPcatNEG2 <- yrSA_CIPcatNEG2 %>% full_join(filter(tmp1, (NCIPp == 2 | NSAp == 2) & `Zp > 0` == FALSE))
  yrSA_CIPcatPOS2 <- yrSA_CIPcatPOS2 %>% full_join(filter(tmp1, (NCIPp == 2 | NSAp == 2) & `Zp > 0` == TRUE))

  yrSA_CIPcatNEG3 <- yrSA_CIPcatNEG3 %>% full_join(filter(tmp1, (NCIPp == 3 | NSAp == 3) & `Zp > 0` == FALSE))
  yrSA_CIPcatPOS3 <- yrSA_CIPcatPOS3 %>% full_join(filter(tmp1, (NCIPp == 3 | NSAp == 3) & `Zp > 0` == TRUE))

  yrSA_CIPcatNEG4 <- yrSA_CIPcatNEG4 %>% full_join(filter(tmp1, (NCIPp == 4 | NSAp == 4) & `Zp > 0` == FALSE))
  yrSA_CIPcatPOS4 <- yrSA_CIPcatPOS4 %>% full_join(filter(tmp1, (NCIPp == 4 | NSAp == 4) & `Zp > 0` == TRUE))

  # }

  yrSA_CIPtotNEG[1,i] <- sum(filter(tmp1, `Zp > 0` == FALSE)[[5]])
  yrSA_CIPtotPOS[1,i] <- sum(filter(tmp1, `Zp > 0` == TRUE)[[5]])

}

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



```

SA_CIPpercTotNEG[4,i] <- sum(filter(yrSA_CIPcatNEG4, Yp == i + 1979)[[5]]) / yrSA_CIPtotNEG[1,i]
SA_CIPpercTotPOS[4,i] <- sum(filter(yrSA_CIPcatPOS4, Yp == i + 1979)[[5]]) / yrSA_CIPtotPOS[1,i]
}

colnames(SA_CIPpercTotNEG) <- c(1980:2018)
colnames(SA_CIPpercTotPOS) <- c(1980:2018)

rownames(SA_CIPpercTotNEG) <- c("SA_CIP1 Zp- %", "SA_CIP2 Zp- %", "SA_CIP3 Zp- %", "SA_CIP4 Zp- %")
rownames(SA_CIPpercTotPOS) <- c("SA_CIP1 Zp+ %", "SA_CIP2 Zp+ %", "SA_CIP3 Zp+ %", "SA_CIP4 Zp+ %")

SA_CIPpercTotPOS

##          1980      1981      1982      1983      1984      1985
## SA_CIP1 Zp+ % 0.9736211 0.959139785 0.96653543 0.97227036 0.97147385 0.9595537
## SA_CIP2 Zp+ % 0.0263789 0.036559140 0.03346457 0.02772964 0.02852615 0.0390516
## SA_CIP3 Zp+ % 0.0000000 0.004301075 0.000000000 0.000000000 0.000000000 0.0013947
## SA_CIP4 Zp+ % 0.0000000 0.000000000 0.000000000 0.000000000 0.000000000 0.0000000
##          1986      1987      1988      1989      1990
## SA_CIP1 Zp+ % 0.95318596 0.951136364 0.942166141 0.934077079 0.933844679
## SA_CIP2 Zp+ % 0.04551365 0.047727273 0.055730810 0.060851927 0.063279003
## SA_CIP3 Zp+ % 0.00130039 0.001136364 0.002103049 0.005070994 0.002876318
## SA_CIP4 Zp+ % 0.00000000 0.000000000 0.000000000 0.000000000 0.000000000 0.0000000
##          1991      1992      1993      1994      1995
## SA_CIP1 Zp+ % 0.948181818 0.918964077 0.910771704 0.89301310 0.89095574
## SA_CIP2 Zp+ % 0.047272727 0.079365079 0.084405145 0.10334789 0.09685696
## SA_CIP3 Zp+ % 0.004545455 0.001670844 0.004823151 0.00363901 0.01218730
## SA_CIP4 Zp+ % 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.0000000
##          1996      1997      1998      1999      2000
## SA_CIP1 Zp+ % 0.8670143416 0.85855072 0.83242059 0.820031299 0.8430992736
## SA_CIP2 Zp+ % 0.1173402868 0.12405797 0.14731654 0.156494523 0.1433414044
## SA_CIP3 Zp+ % 0.0149934811 0.01565217 0.01916758 0.022430882 0.0130750605
## SA_CIP4 Zp+ % 0.0006518905 0.00173913 0.00109529 0.001043297 0.0004842615
##          2001      2002      2003      2004      2005
## SA_CIP1 Zp+ % 0.808633734 0.819243156 0.796847008 0.800247372 0.783571221
## SA_CIP2 Zp+ % 0.171339564 0.161030596 0.174131136 0.169449598 0.185843286
## SA_CIP3 Zp+ % 0.018691589 0.018518519 0.027230383 0.028138528 0.029129042
## SA_CIP4 Zp+ % 0.001335113 0.001207729 0.001791473 0.002164502 0.001456452
##          2006      2007      2008      2009      2010
## SA_CIP1 Zp+ % 0.777632268 0.75808056 0.745620182 0.748487468 0.721241761
## SA_CIP2 Zp+ % 0.191985333 0.20313277 0.209296893 0.206352636 0.229002764
## SA_CIP3 Zp+ % 0.029334730 0.03580308 0.042046251 0.039974071 0.046353391
## SA_CIP4 Zp+ % 0.001047669 0.00298359 0.003036674 0.005185825 0.003402084
##          2011      2012      2013      2014      2015
## SA_CIP1 Zp+ % 0.722359227 0.704897959 0.705371567 0.697417743 0.686005314
## SA_CIP2 Zp+ % 0.229346486 0.241020408 0.243336026 0.254060808 0.258193091
## SA_CIP3 Zp+ % 0.044184135 0.049183673 0.044830372 0.043315285 0.050487157
## SA_CIP4 Zp+ % 0.004110152 0.004897959 0.006462036 0.005206164 0.005314438
##          2016      2017      2018
## SA_CIP1 Zp+ % 0.688063311 0.726166329 0.735719726
## SA_CIP2 Zp+ % 0.259177841 0.224898580 0.218075654
## SA_CIP3 Zp+ % 0.047922620 0.043103448 0.041127190
## SA_CIP4 Zp+ % 0.004836228 0.005831643 0.005077431

```

```
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=p3, mapping = aes(x=years,y=Fraction), size=0.5, color="royalblue") + geom_line(data=n1, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey30") + geom_line(data=n3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="royalblue")  
  
e2<-e1+theme_bw() + theme(panel.border = element_rect(colour = "black", fill=NA, size=1), panel.grid.major = element_rect(colour = "black", fill=NA, size=1), panel.grid.minor = element_rect(colour = "black", fill=NA, size=0.5), 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, 100))  
  
e4<-e3+labs(title = "", x = "", y = "")  
  
e5<-e4+ scale_color_manual(name = "Y series")
```

#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=cp3, mapping = aes(x=years,y=Fraction), size=0.5, color="royalblue") + geom_line(data=cn1, mapping = aes(x=years,y=Fraction), size=0.5, color="grey30") + geom_line(data=cn3, mapping = aes(x=years,y=Fraction), size=0.5, color="royalblue")
```

```

geom_line(data=cn1, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey30")
geom_line(data=cn3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="royalblue")

ce2<-ce1+theme_bw() + theme(panel.border = element_rect(colour = "black", fill=NA, size=1), 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(0.23, 0.8))

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(1983, 2007, 1)), x=1983, y=0.23))

## 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).

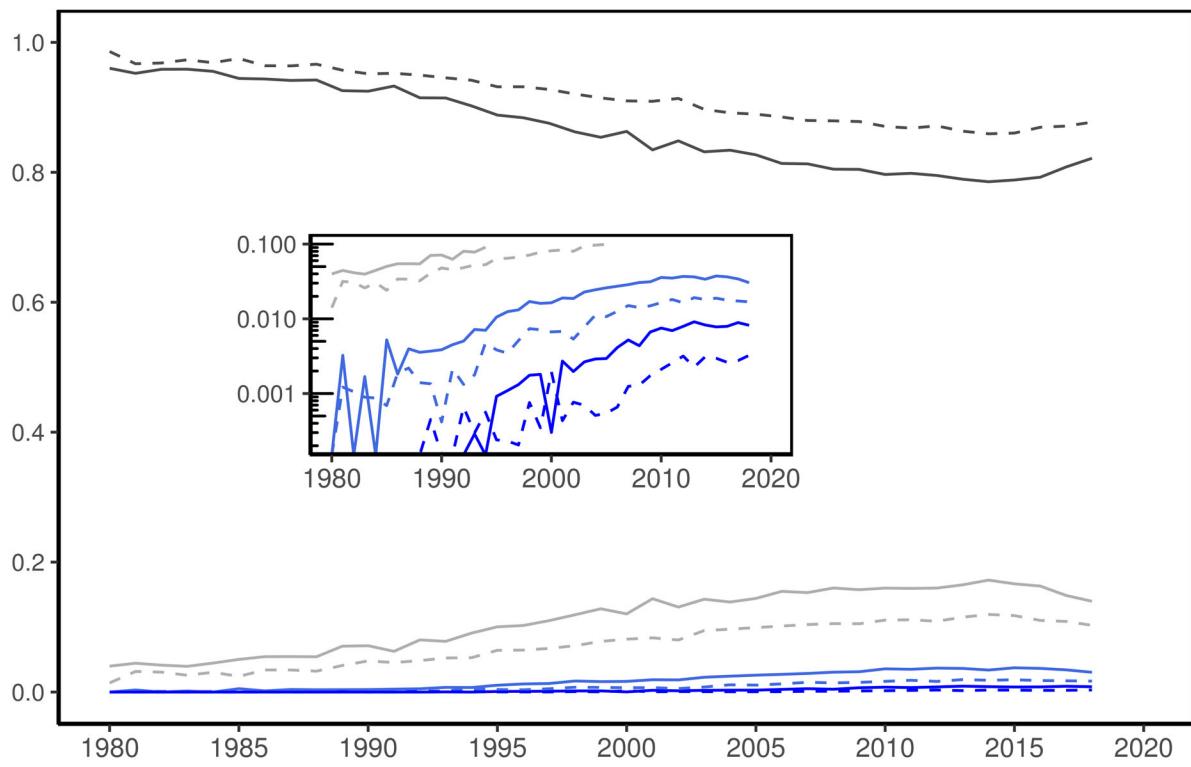
## Warning: Removed 24 row(s) containing missing values (geom_path).

## Warning: Removed 39 row(s) containing missing values (geom_path).

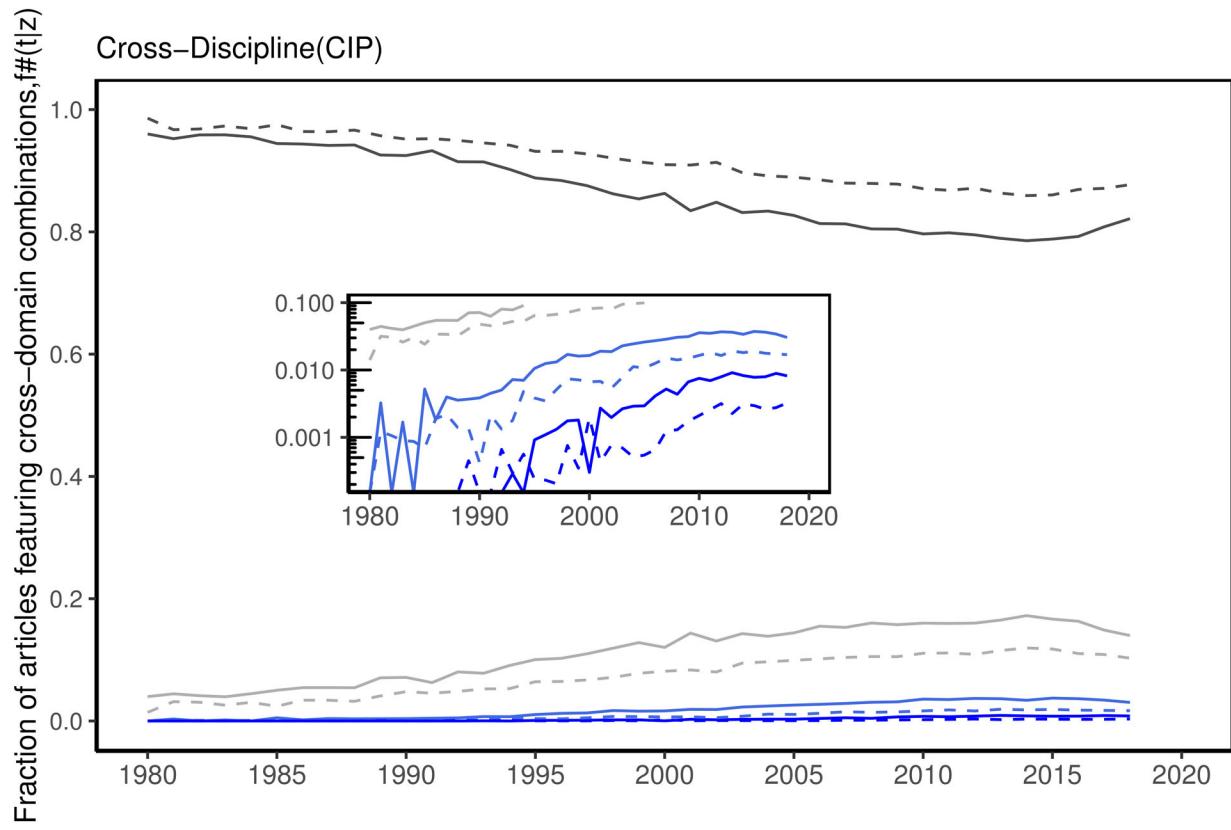
## Warning: Removed 13 row(s) containing missing values (geom_path).

ce5

```



```
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)")+ylab("Fraction of articles featuring cross-domain combinations")
```



```
#Plotting for SACIP
```

```

scp1 = data.frame(years=years, Fraction = (SA_CIPpercTotPOS[1,]))
scp2 = data.frame(years=years, Fraction = (SA_CIPpercTotPOS[2,]))
scp3 = data.frame(years=years, Fraction = SA_CIPpercTotPOS[3,])
scp4 = data.frame(years=years, Fraction = (SA_CIPpercTotPOS[4,]))


scn1 = data.frame(years=years, Fraction = (SA_CIPpercTotNEG[1,]))
scn2 = data.frame(years=years, Fraction = (SA_CIPpercTotNEG[2,]))
scn3 = data.frame(years=years, Fraction = SA_CIPpercTotNEG[3,])
scn4 = data.frame(years=years, Fraction = (SA_CIPpercTotNEG[4,]))


sce1<-ggplot() +geom_line(data=scp1, mapping = aes(x=years,y=Fraction), size=0.5, color="grey30") + geom_
  geom_line(data=scp3, mapping = aes(x=years,y=Fraction), size=1, color="royalblue") + geom_line(data=scn1,
  geom_line(data=scn3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="grey30")
  geom_line(data=scn3, mapping = aes(x=years,y=Fraction), linetype = "dashed", size=0.5, color="royalbl

sce2<-sce1+theme_bw() + theme(panel.border = element_rect(colour = "black", fill=NA, size=1), panel.grid_
  panel.grid.minor = element_blank(), axis.line = element_line(colour = "black")
  axis.text.x   = element_text(size=10))
sce3<-sce2+scale_x_continuous(limits=c(1980, 2020), breaks=seq(1980, 2020, 5))+scale_y_continuous(limits=c(0, 1))

sce4<-sce3+labs(title = "", x = "", y = "")
```

```

sce5<-sce4 +annotation_custom(
  ggplotGrob(sce4+annotation_logticks() +
    scale_y_log10(limits = c(NA,0.1))+scale_x_continuous(limits=c(1980, 2020), breaks=seq(1980, 2005, 20, 2020)),
    xmin = 1980, xmax = 2005, ymin = 0.19, ymax = 0.79)

## 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

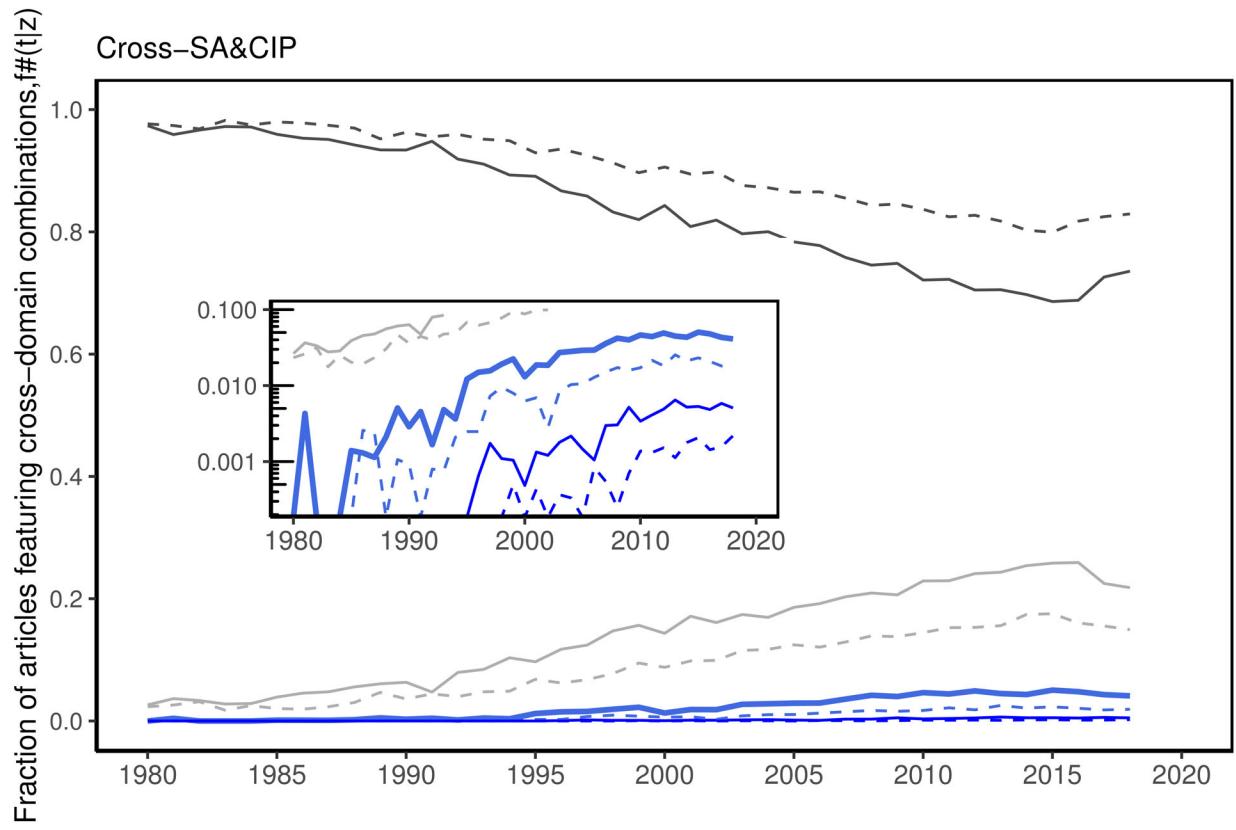
## Warning: Removed 39 row(s) containing missing values (geom_path).

## Warning: Removed 23 row(s) containing missing values (geom_path).

## Warning: Removed 39 row(s) containing missing values (geom_path).

## Warning: Removed 16 row(s) containing missing values (geom_path).

sce5+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-SA&CIP")+ylab("Fraction of articles featuring cross-domain combinations,f#(t|z)"
```



```
grid.arrange(arrangeGrob(e5 + theme(legend.position="none"),
  ce5 + theme(legend.position="none"),
  sce5 + theme(legend.position="none"),
  nrow = 3,
  top = textGrob("#Categories", vjust = 1, gp = gpar(fontface = "bold", cex = 1),
  left = textGrob("Fraction of articles featuring cross-domain combinations,f#(t|z)", fontface = "bold", cex = 1),
  right = textGrob("Cross-Topic(SA) Cross-Discipline(CIP)", fontface = "bold", cex = 1))
```

