Project Milestone II Figure 3A

Ryan Thackston

4/8/2021

```r
library(reshape2)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```r
library(grid)
setwd("G:\\COSC_6323_Statistics_for_Researchers\\Project")
df1<-read.csv("ArticleLevel-RegData-ALLSA_Xc_1_NData_655386_LONGXCIP2.csv")

# Group by Years
df1 %>%
  mutate(year_group = case_when(
    (Yp %in% c(1970,1971,1972,1973,1974)) ~ "1970-1974",
    (Yp %in% c(1975,1976,1977,1978,1979)) ~ "1975-1979",
    (Yp %in% c(1980,1981,1982,1983,1984)) ~ "1980-1984",
    (Yp %in% c(1985,1986,1987,1988,1989)) ~ "1985-1989",
    (Yp %in% c(1990,1991,1992,1993,1994)) ~ "1990-1994",
    (Yp %in% c(1995,1996,1997,1998,1999)) ~ "1995-1999",
    (Yp %in% c(2000,2001,2002,2003,2004)) ~ "2000-2004",
    (Yp %in% c(2005,2006,2007,2008,2009)) ~ "2005-2009",
    (Yp %in% c(2010,2011,2012,2013,2014)) ~ "2010-2014",
    (Yp %in% c(2015,2016,2017,2018))      ~ "2015-2018",
  )) -> dfyg
```

```r
# Removing the NA values from the year group
#plots <- list()


# Need to make a nested tibble so i can call on index to create a new tibble
# Or I need to make a 3D data frame, store each df_CIP1_grp, then convert them to tibbles.

my_list <- vector(mode = "list", length = 9)

for (i in  c(1:9)){
  # Filters out only papers with certain CIP# index
  # Is a data frame
  df_CIP1 <- subset(dfyg, dfyg[17+i] >0)

  # Prints CIP column #
  print(colnames(dfyg)[17+i])


  # after running this line, the data frame gets saved as a tibble
  df_CIP1 %>%

    group_by(year_group) %>%
    summarise(
      SA1 = sum(SA1 * df_CIP1[17+i]),
      SA2 = sum(SA2 * df_CIP1[17+i]),
      SA3 = sum(SA3 * df_CIP1[17+i]),
      SA4 = sum(SA4 * df_CIP1[17+i]),
      SA5 = sum(SA5 * df_CIP1[17+i]),
      SA6 = sum(SA6 * df_CIP1[17+i])

    ) -> df_CIP1_grp


df_CIP1_grp$SA_sum <- df_CIP1_grp$SA1 + df_CIP1_grp$SA2 + df_CIP1_grp$SA3+
  df_CIP1_grp$SA4 + df_CIP1_grp$SA5 + df_CIP1_grp$SA6

df_CIP1_grp$SA1_Fraction<-df_CIP1_grp$SA1/df_CIP1_grp$SA_sum
df_CIP1_grp$SA2_Fraction<-df_CIP1_grp$SA2/df_CIP1_grp$SA_sum
df_CIP1_grp$SA3_Fraction<-df_CIP1_grp$SA3/df_CIP1_grp$SA_sum
df_CIP1_grp$SA4_Fraction<-df_CIP1_grp$SA4/df_CIP1_grp$SA_sum
df_CIP1_grp$SA5_Fraction<-df_CIP1_grp$SA5/df_CIP1_grp$SA_sum
df_CIP1_grp$SA6_Fraction<-df_CIP1_grp$SA6/df_CIP1_grp$SA_sum


df_CIP1_grp$SA_sum<-NULL

df_CIP1_grp<-df_CIP1_grp[-11,]

# Rename Columns from SA to actual name
colnames(df_CIP1_grp)<-c("year_group",
                         "P&P",
                         "A&O",
                         "Ph&Pr",
```

```r
                             "H",
                             "T&E",
                             "T&IS",
                             "P&P_Frac",
                             "A&O_Frac",
                             "Ph&Pr_Frac",
                             "H_Frac",
                             "T&E_Frac",
                             "T&IS_Frac")

my_list[[i]] <- df_CIP1_grp

# ylabs = c("Neurosciences", "Biology", "Psychology",
#           "Biotech & Genetics", "Medical Specialty", "Health Sciences",
#           "Pathology & Pharmacology", "Eng. & Informatics",
#           "Chem. & Physics & Math")

ylabs = c("Neurosciences","Biotech & Genetics","Medical Specialty",
          "Eng. & Informatics", "Biology", "Psychology",
          "Health Sciences","Pathology & Pharmacology",
          "Chem & Phy & Math")

}
```

```
## [1] "CIP1"
## [1] "CIP2"
## [1] "CIP3"
## [1] "CIP4"
## [1] "CIP5"
## [1] "CIP6"
## [1] "CIP7"
## [1] "CIP8"
## [1] "CIP9"
```

```r
plot_list <- vector(mode = "list", length = 9)

  for (i in  c(1:9) ) {
    mlt_df_CIP1_grp <-melt(my_list[[i]], id="year_group")

    f <- mlt_df_CIP1_grp[61:120,]
    # colnames(f)[colnames(f)=="value"]<- ylabs[i]

    gg_plot<-ggplot(f)
    # gg_plot1<-gg_plot + geom_bar(aes(x=year_group, y=ylabs[i], fill=forcats::fct_rev(variable)),
    #                              stat="identity")
    gg_plot1<-gg_plot + geom_bar(aes(x=year_group, y=value, fill=forcats::fct_rev(variable)),
                                 stat="identity")



    gg_plot12<-gg_plot1+theme(axis.text.x = element_text(angle = 45, hjust = 1),
                              axis.title.x=element_blank(),legend.position="none",
                              axis.text.y = element_text(angle = 90, hjust= 1 ))
```

```r
    plot_list[[i]]<-gg_plot12+ylab(colnames(df_CIP1[17+i])+scale_fill_discrete(name="Subject Areas: "))
      scale_fill_manual(values = rev(c("#cc0000","#ff9966","#91eb83","darkgreen","black","darkgrey")))

    plot_list[[i]]<- plot_list[[i]] + labs(y = ylabs[i])

    if(i %in% c(1,2,3,4,5,8)){
      plot_list[[i]]<-plot_list[[i]]+theme(axis.text.x = element_blank())
    }

  }

ggarrange(
  plot_list[[1]],
  plot_list[[4]],
  plot_list[[5]],
  plot_list[[8]],
  plot_list[[2]],
  plot_list[[3]],
  plot_list[[6]],
  plot_list[[7]],
  plot_list[[9]],
  ncol = 3,
  nrow = 3,
  labels = NULL,
  label.x = 0,
  label.y = 1,
  hjust = -0.5,
  vjust = 1.5,
  font.label = list(size = 14, color = "black", face = "bold", family = NULL),
  align = c("none", "h", "v", "hv"),
  widths = 1,
  heights = 1,
  legend = NULL,
  common.legend = FALSE,
  legend.grob = NULL
)
```