WEEK 12 NOV 10-NOV 16 lab 9 Population Aging Using % in age-sex group-faceted charts

Lysa Vanible

November 10, 2021

#**Introduction**  
In this lab we analyze population aging. We will use the R’s ggplot package to construct population pyramids for the USA and Japan. Population pyramids are bar charts that show the number or % of individuals in each age-sex group of a population. They are used for examining population aging as well as if a population is growing, stable, or shrinking in size. The pyramids we will construct in this lab will show the % of people in each age sex group. You will need the username and password for mortality.org that you created during lab 2.

1). First replace “Write your name here” in the YAML header above with your name before you start. MAKE SURE NOT TO MODIFY ANY OTHER PART OF THE HEADER.

2). Again the same general rules used in the previous labs hold for this lab. Do not delete or modify anything unless it was written by you or you are asked to do so. All R code must be entered in the grey boxes/code chunks.

3). To run R code as you work your way through the assignment, you can highlight the code and click on the “Run” command in the R studio menu and then on “Run Selected Lines” or click on the green button at the top right of a particular code chunk to run all the code within it.

4). When you get to the end of the assignment you will knit the document to produce a pdf document using the Knit command in the Rstudio menu.

5). Continue reading from Part A.

knitr::opts\_chunk$set(error = TRUE)

PART A: CREATING POPULATION PYRAMIDS FOR USA Let us start by loading the relevant libraries.

library(HMDHFDplus)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.1.2 ✓ dplyr 1.0.6  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

We’ll use data from mortality.org for this lab. First, let’s get the population data for the USA. Enter your username and password for mortality.org in the appropriate places in the code below and run the code. The code will extract the male and female populations in single age groups from 1933 for the USA. It will then construct a new variable AgeGroup made of 5 year age groups 0-4, 5-9 etc and compute the male and populations in those 5 year age groups.

population\_USA <- readHMDweb(CNTRY = 'USA', item = 'Population',  
username='lysa.vanible@lc.cuny.edu', password='Oceansky2004')  
population\_USA<-select(population\_USA,Year,Age,Female1,Male1)  
colnames(population\_USA)<-c("Year","Age","Female","Male")  
population\_USA\_1<-pivot\_longer(population\_USA, names\_to = 'Gender', values\_to = 'Population', cols=c(3:4))  
labs <- c(paste(seq(0, 95, by = 5), seq(0 + 5 - 1, 100 - 1, by = 5),  
 sep = "-"), paste(100, "+", sep = ""))  
population\_USA\_1$AgeGroup <- cut(population\_USA\_1$Age, breaks = c(seq(0, 100, by = 5), Inf), labels = labs, right = FALSE)  
population\_USA\_1<-select(population\_USA\_1, Year,AgeGroup,Gender,Population)  
population\_USA\_2<-aggregate(formula = Population ~Year+Gender+AgeGroup, data = population\_USA\_1, FUN = sum)

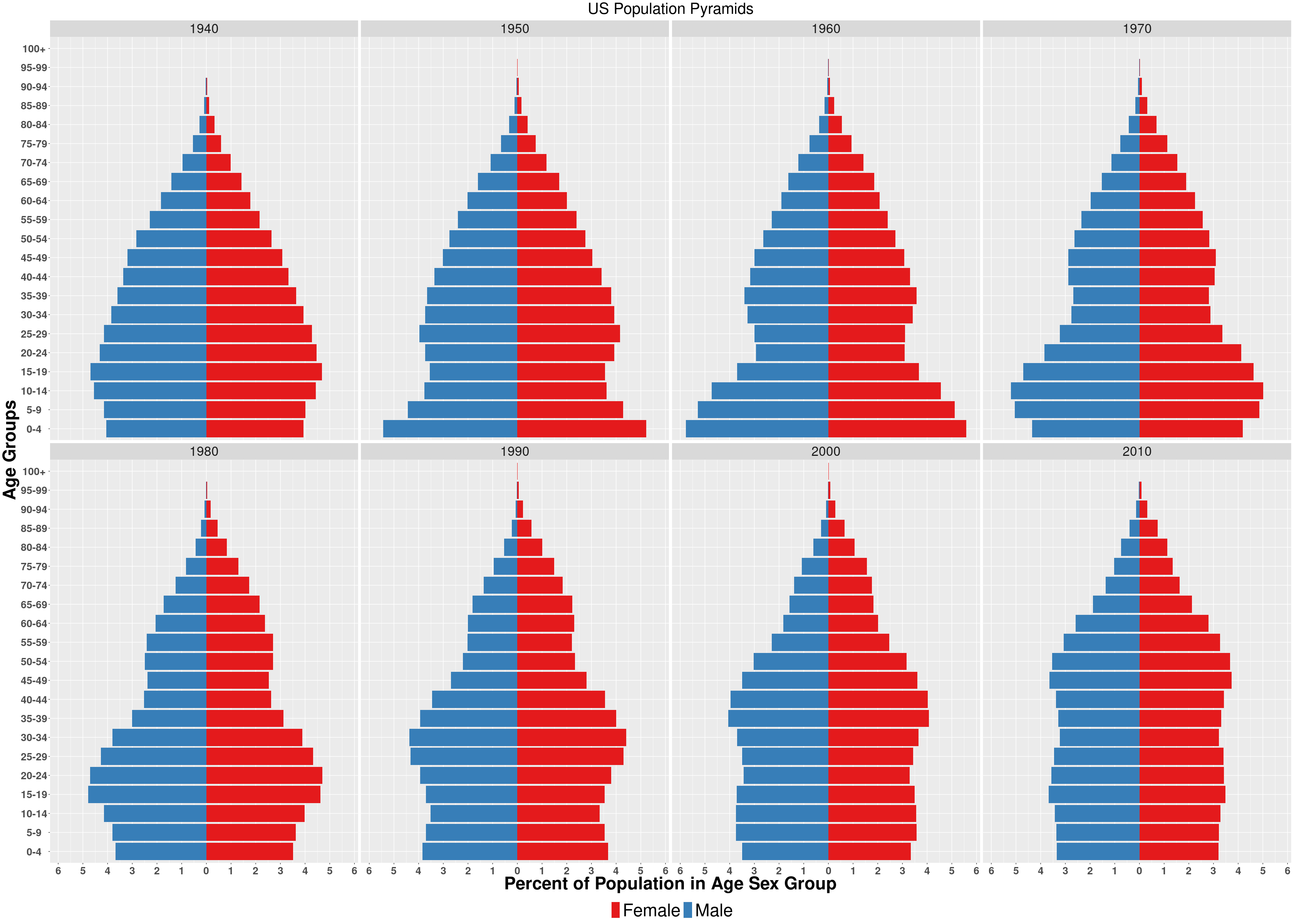
The code below gathers the population data for the last year of each decade into variables starting from 1940. For instance, usa\_filtered1940 included 1940 population data etc. Data for all the other years are discarded. usa\_filtered1950 included 1950 population data etc and discarded data all the other years. The code also computes the percentage of the population in each age-sex group for each each year’s population.

usa\_filtered1940 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1940)  
usa\_filtered1940\_4<-usa\_filtered1940%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered1950 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1950)  
usa\_filtered1950\_4<-usa\_filtered1950%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered1960 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1960)  
usa\_filtered1960\_4<-usa\_filtered1960%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered1970 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1970)  
usa\_filtered1970\_4<-usa\_filtered1970%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered1980 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1980)  
usa\_filtered1980\_4<-usa\_filtered1980%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered1990 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1990)  
usa\_filtered1990\_4<-usa\_filtered1990%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered2000 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==2000)  
usa\_filtered2000\_4<-usa\_filtered2000%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
usa\_filtered2010 <- select(population\_USA\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==2010)  
usa\_filtered2010\_4<-usa\_filtered2010%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))

The code below places the age-sex percentages for the years, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, together in the variable total\_USA. Then the ggplot code below uses the data on the percentages of the population in each age-sex group for each year to construct the population pyramids for each year. The pyramids are placed next to each other to make comparison easy and stored in an object named “u”. Please run the code.

total\_USA<-rbind(usa\_filtered1940\_4,usa\_filtered1950\_4,usa\_filtered1960\_4,usa\_filtered1970\_4,usa\_filtered1980\_4, usa\_filtered1990\_4,usa\_filtered2000\_4,usa\_filtered2010\_4)  
  
u<- ggplot(total\_USA, aes(x = AgeGroup, y = PopPerc, fill = Gender)) +   
 geom\_bar(data = subset(total\_USA, Gender == "Female"), stat = "identity") +  
 geom\_bar(data = subset(total\_USA, Gender == "Male"), stat = "identity") +   
 scale\_y\_continuous(breaks=seq(-6,6,1),labels=abs(seq(-6,6,1)))+  
 coord\_flip()+  
 scale\_fill\_brewer(palette = "Set1") +   
 facet\_wrap(~Year, nrow=2)+  
 labs(x = "Age Groups", y = "Percent of Population in Age Sex Group", fill = "",   
 title = "US Population Pyramids")+  
 theme(axis.text.x = element\_text(angle = 0, hjust = 0.5,   
 vjust = 0.5, size=20,face="bold"),plot.title = element\_text(hjust = 0.5,size=30),axis.text.y = element\_text(angle = 0, hjust = 0.5, vjust = 0.5, size=20,face="bold"), strip.text.x = element\_text(  
 size = 25),legend.text =element\_text(size=34),legend.position="bottom",axis.title=element\_text(size=34,face="bold"))

Run the code below to see the population pyramids placed in the object “u”.

 PART B: CREATING POPULATION PYRAMIDS FOR JAPAN

Now let us create population pyramids for Japan. Once again, fill in your mortality.org username and password in the code below. Then go ahead and run the code below to extract Japan population data for the various years and get them ready to compute the % of the population in age-sex group.

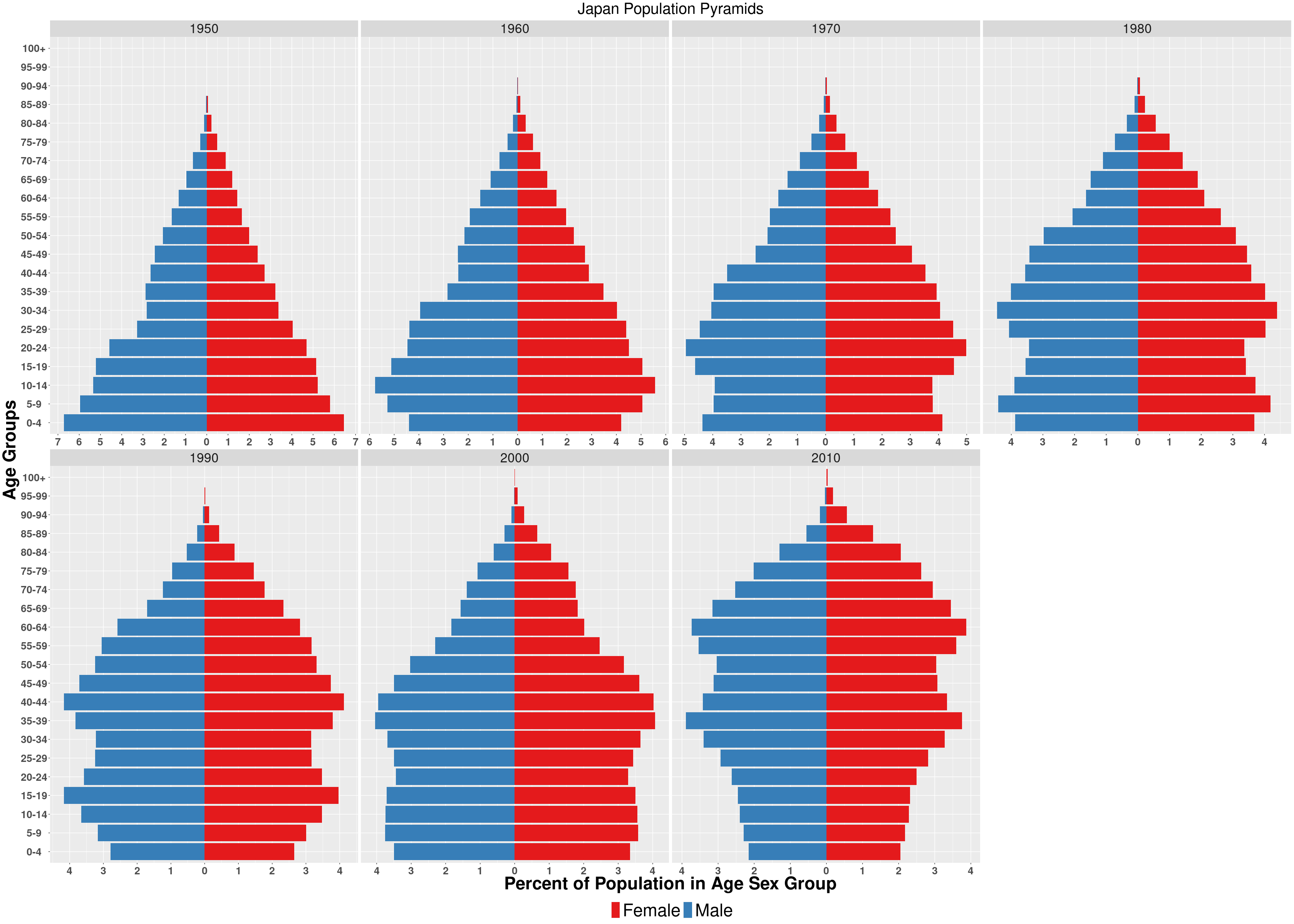
population\_JPN <- readHMDweb(CNTRY = 'JPN', item = 'Population',  
username='lysa.vanible@lc.cuny.edu', password='Oceansky2004')  
population\_JPN<-select(population\_JPN,Year,Age,Female1,Male1)  
colnames(population\_JPN)<-c("Year","Age","Female","Male")  
population\_JPN\_1<-pivot\_longer(population\_JPN, names\_to = 'Gender', values\_to = 'Population', cols=c(3:4))  
labs <- c(paste(seq(0, 95, by = 5), seq(0 + 5 - 1, 100 - 1, by = 5),  
 sep = "-"), paste(100, "+", sep = ""))  
population\_JPN\_1$AgeGroup <- cut(population\_JPN\_1$Age, breaks = c(seq(0, 100, by = 5), Inf), labels = labs, right = FALSE)  
population\_JPN\_1<-select(population\_JPN\_1, Year,AgeGroup,Gender,Population)  
population\_JPN\_2<-aggregate(formula = Population ~Year+Gender+AgeGroup, data = population\_JPN\_1, FUN = sum)

The code here computes the percentages of the Japanese population in each age-sex group for last year of each decade starting from 1950. Go ahead and run the code.

JPN\_filtered1950<- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1950)  
JPN\_filtered1950\_4<-JPN\_filtered1950%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered1960 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1960)  
JPN\_filtered1960\_4<-JPN\_filtered1960%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered1970 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1970)  
JPN\_filtered1970\_4<-JPN\_filtered1970%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered1980 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1980)  
JPN\_filtered1980\_4<-JPN\_filtered1980%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered1990 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==1990)  
JPN\_filtered1990\_4<-JPN\_filtered1990%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered2000 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==2000)  
JPN\_filtered2000\_4<-usa\_filtered2000%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))  
  
JPN\_filtered2010 <- select(population\_JPN\_2,Year,AgeGroup,Gender,Population)%>%filter(Year==2010)  
JPN\_filtered2010\_4<-JPN\_filtered2010%>%mutate(PopPerc=if\_else(Gender=="Female", round(Population/sum(Population)\*100,2),-round(Population/sum(Population)\*100,2)))

The code below places all the age-sex percentages constructed in the preceding together in total\_JPN. The ggplot code then uses the data on the percentages of the population in each age-sex group for each year to construct the population pyramids for Japan. Once again, the pyramids are placed next to each other so it is relatively easy to compare them and see how they are changing over time. The pyramids are placed in an object named “j”.

total\_JPN<-rbind(JPN\_filtered1950\_4,JPN\_filtered1960\_4,JPN\_filtered1970\_4,JPN\_filtered1980\_4, JPN\_filtered1990\_4,JPN\_filtered2000\_4,JPN\_filtered2010\_4)  
  
j<- ggplot(total\_JPN, aes(x = AgeGroup, y = PopPerc, fill = Gender)) +   
 geom\_bar(data = subset(total\_JPN, Gender == "Female"), stat = "identity") +  
 geom\_bar(data = subset(total\_JPN, Gender == "Male"), stat = "identity") +   
 scale\_y\_continuous(breaks=seq(-7,7,1),labels=abs(seq(-7,7,1)))+  
 coord\_flip()+  
 scale\_fill\_brewer(palette = "Set1") +   
 facet\_wrap(~Year, scales="free\_x", nrow=2)+  
 labs(x = "Age Groups", y = "Percent of Population in Age Sex Group", fill = "",   
 title = "Japan Population Pyramids")+  
 theme(axis.text.x = element\_text(angle = 0, hjust = 0.5,   
 vjust = 0.5, size=20, face="bold"),plot.title = element\_text(hjust = 0.5,size=30),axis.text.y = element\_text(angle = 0, hjust = 0.5, vjust = 0.5, size=20, face="bold"), strip.text.x = element\_text(  
 size = 25),legend.text =element\_text(size=34),legend.position="bottom",axis.title=element\_text(size=34,face="bold"))

Now run the code below to see Japan’s pyramids stored in the object “j”.  PART C CONCLUSION

Great work!!!! You are almost at the end of your ninth lab. You will now prepare to knit the lab. Click on the “Knit” command at the top of this window to turn this document with your code and all the results into a pdf document. Inspect the knitted document very carefully to make sure that you have not missed any part of the assignment. If you have missed any tasks please go back and fix them and knit the document again. Once the pdf document is knitted and you are sure you have completed all the assignments in it prepare to submit it to the week 12 folder of SOC339’s blackboard site. To do that there are a few more things to do.

1). First save your rmarkdown file (this the file you have been working on in Rstudio) by clicking on the FILE command in the Rstudio menu and then on SAVE. You will see the rmarkdown document in the FILES folder of the window to the bottom right of your screen. It will be named Lab 9 Population Aging.Rmd.

2). To save the pdf file that you produced, locate it in the FILES folder of the window to the bottom right of your screen. It will be named Lab 9 Population Aging.pdf. Click the little box to the left of it so a check mark appears in it. Then click on the MORE tab near the top of the FILES window. Then on EXPORT and then when prompted by RSTUDIO click on DOWNLOAD. RSTUDIO will save the lab 9 Population Aging.pdf to the DOWNLOAD folder of your computer.

4). Go to the WEEK 12 FOLDER in blackboard and follow the instructions inside that folder to upload Lab 9 Population Aging.pdf and complete the remaining questions in the lab.