COS-R403. Special Research Methods. *Forecasting I: Introduction*

Hands-on excercises

Day 1 of intensive 5-day course

University of Helsinki, Finland

04.05.2020–08.05.2020

Lecturer: Christina Bohk-Ewald

Source: https://github.com/christina-bohk-ewald/2020-COS-R403-forecasting-I-introduction

Table of content:

Feel free to change and adapt R code to your needs and creativity.

1. Some preparations in R

1.1 Open a new script for this hands-on exercise in R and save it to a folder of your choice.

1.2 Create a filepath to a folder where you would like to save your input data. For example,

```
the.data.path <- c("C:/data")
```

1.3 Create a filepath to a folder where you would like to save your outcome. For example,

```
the.plot.path <- c("C:/plots")
```

1.4 You can then set the working directory to this input data path or outcome path

```
setwd(the.data.path)

setwd(the.plot.path)
```

2. Download and save UNWPP2019 data from website to load and explore them in R

2.1 Please go to the UNWPP2019 website and explore the material that is available. Download data files that provide

- Global population counts for females and males
- Global life expectancy at birth for females and males
- Global total fertility rate
- Probabilistic forecasts of global population counts
- Probabilistic forecasts of global life expectancy at birth for females and males
- Probabilistic forecasts of global total fertility rate

and save them in your created folder for input data (section 1).

2.2 Load these input data in R. Start with observed data, then continue with forecasted data.

```
setwd(the.data.path)
```

```
## install.packages("openxlsx")
require(openxlsx)

####
```

```
##
## 1. Load observed data
##
####

##
### 1.1 Population counts
##

wom <- read.xlsx(file.path(the.data.path,
paste("WPP2019_INT_F03_3_POPULATION_BY_AGE_ANNUAL_FEMALE.xlsx",sep="")),
sheet = 1,startRow = 17)
## dim(wom)
## wom[1:2,1:10]
## wom[1:2,c(6,8,9,10)]
wom_select <- wom[,c(3,8:109)]
wom_select[1:2,1:5]
```

```
##   Region,.subregion,.country.or.area.* Reference.date.(as.of.1.July)
## 1                                WORLD                          1950
## 2                                WORLD                          1951
##                    0         1                 2
## 1           38581.47 35347.945          32693.216
## 2 40594.885999999999 37655.267 34738.559000000001
```

```
## unique(wom_select[,1])
```

Please explain the function *read.xlsx()*.

Brief data description. *wom_select* contains population counts by region (column 1), calendar year (column 2), and ages 0 through 100 (columns 3 through 103). With *unique(wom_select[,1])* you can list the regions for which population counts are available.

Please continue loading input data in R:

```
men <- read.xlsx(file.path(the.data.path,
paste("WPP2019_INT_F03_2_POPULATION_BY_AGE_ANNUAL_MALE.xlsx",sep="")),
sheet = 1,startRow = 17)
men_select <- men[,c(3,8:109)]
men_select[1:2,1:5]
```

```
##   Region,.subregion,.country.or.area.* Reference.date.(as.of.1.July)
## 1                                WORLD                          1950
## 2                                WORLD                          1951
##                    0                 1                 2
## 1          39854.036          36612.375          33956.483
## 2 42186.656999999999 38929.764000000003 35995.798999999999
```

```
##
### 1.2 Life expectancy at birth
##

e0_wom <- read.xlsx(file.path(the.data.path,
paste("WPP2019_MORT_F07_3_LIFE_EXPECTANCY_0_FEMALE.xlsx",sep="")),
sheet = 1,startRow = 17)
e0_wom_select <- e0_wom[1,8:21]
e0_wom_select_world_regions <- e0_wom[18:23,8:21]
```

```
e0_wom_select_world_regions[1:2,1:5]
```

```
##             1950-1955          1955-1960          1960-1965          1965-1970
## 18 38.746865415063802 41.356266858619499 43.808289407637801 45.924243355210102
## 19 43.200750154383499 45.554491389260399   47.4337870739054   53.8729789847864
##           1970-1975
## 18 48.0692902052104
## 19 57.4947590287209
```

```
e0_men <- read.xlsx(file.path(the.data.path,
paste("WPP2019_MORT_F07_2_LIFE_EXPECTANCY_0_MALE.xlsx",sep="")),
sheet = 1,startRow = 17)
e0_men_select <- e0_men[1,8:21]
e0_men_select_world_regions <- e0_men[18:23,8:21]
e0_men_select_world_regions[1:2,1:5]
```

```
##             1950-1955          1955-1960          1960-1965          1965-1970
## 18 36.281840336756602 38.759864506767798 41.125958926145998 43.176137034902503
## 19 41.480359419528199 43.672643918754702 45.617716547674199 51.544308899177103
##             1970-1975
## 18 45.194186649810803
## 19   55.3101089935152
```

Brief data description. *e0_wom_select* contains global life expectancy at birth (e0) for calendar years 1950-1955 through 2015-2020. Life expectancy at birth of other world regions is stored in *e0_wom_select_world_regions*. Please note that it is up to you to find a proper way to organize your data.

```
##
### 1.3 Total fertility rate
##

tfr <- read.xlsx(file.path(the.data.path,
paste("WPP2019_FERT_F04_TOTAL_FERTILITY.xlsx",sep="")),sheet = 1,startRow = 17)
tfr_select <- tfr[1,8:21]
tfr_select[1,1:5]
```

```
##            1950-1955          1955-1960          1960-1965        1965-1970
## 1 4.9673756888117504 4.8972484324963697 5.0181079377603899 4.92634162451915
##            1970-1975
## 1 4.4711311113813599
```

Brief data description. *tfr_select* contains global total fertility rate (TFR) for calendar years 1950-1955 through 2015-2020. How do you want to save TFR of other world regions?

2.3 Continue loading forecasted data in R following the same procedure. This time you need to insert the excel sheet number yourself in function read.xlsx().

```
####
##
## 2. Load forecasted data
##
####

##
### 2.1.1 Population counts
```

```
##

wom_2 <- read.xlsx(file.path(the.data.path,
paste("WPP2019_INT_F03_3_POPULATION_BY_AGE_ANNUAL_FEMALE.xlsx",sep="")),
sheet = params$sheet_forecast,startRow = 17)
wom_select_2 <- wom_2[,c(3,8:109)]

men_2 <- read.xlsx(file.path(the.data.path,
paste("WPP2019_INT_F03_2_POPULATION_BY_AGE_ANNUAL_MALE.xlsx",sep="")),
sheet = params$sheet_forecast,startRow = 17)
men_select_2 <- men_2[,c(3,8:109)]

##
### 2.1.2 Population counts (median, 80% and 95% prediction intervals)
##

total_pop <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Output_PopTot.xlsx",sep="")),sheet = 3,startRow = 17)
total_pop_median <- total_pop[1,8:24]

total_pop <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Output_PopTot.xlsx",sep="")),sheet = 2,startRow = 17)
total_pop_low_80 <- total_pop[1,8:24]

total_pop <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Output_PopTot.xlsx",sep="")),sheet = 4,startRow = 17)
total_pop_up_80 <- total_pop[1,8:24]

total_pop <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Output_PopTot.xlsx",sep="")),sheet = 1,startRow = 17)
total_pop_low_95 <- total_pop[1,8:24]

total_pop <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Output_PopTot.xlsx",sep="")),sheet = 5,startRow = 17)
total_pop_up_95 <- total_pop[1,8:24]

##
### 2.2.1 Life expectancy at birth
##

e0_wom_2 <- read.xlsx(file.path(the.data.path,
paste("WPP2019_MORT_F07_3_LIFE_EXPECTANCY_0_FEMALE.xlsx",sep="")),
sheet = params$sheet_forecast,startRow = 17)
e0_wom_2_select <- e0_wom_2[1,8:23]
e0_wom_2_select_world_regions <- e0_wom_2[18:23,8:23]

e0_men_2 <- read.xlsx(file.path(the.data.path,
paste("WPP2019_MORT_F07_2_LIFE_EXPECTANCY_0_MALE.xlsx",sep="")),
sheet = params$sheet_forecast,startRow = 17)
e0_men_2_select <- e0_men_2[1,8:23]
e0_men_2_select_world_regions <- e0_men_2[18:23,8:23]

##
```

```
### 2.2.2 Life expectancy at birth (median, 80% and 95% prediction intervals)
##

e0_men_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0M.xlsx",sep="")),sheet = 3,startRow = 17)
e0_men_prob_median <- e0_men_prob[1,8:23]

e0_men_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0M.xlsx",sep="")),sheet = 2,startRow = 17)
e0_men_prob_low_80 <- e0_men_prob[1,8:23]

e0_men_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0M.xlsx",sep="")),sheet = 4,startRow = 17)
e0_men_prob_up_80 <- e0_men_prob[1,8:23]

e0_men_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0M.xlsx",sep="")),sheet = 1,startRow = 17)
e0_men_prob_low_95 <- e0_men_prob[1,8:23]

e0_men_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0M.xlsx",sep="")),sheet = 5,startRow = 17)
e0_men_prob_up_95 <- e0_men_prob[1,8:23]

##

e0_wom_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0F.xlsx",sep="")),sheet = 3,startRow = 17)
e0_wom_prob_median <- e0_wom_prob[1,8:23]

e0_wom_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0F.xlsx",sep="")),sheet = 2,startRow = 17)
e0_wom_prob_low_80 <- e0_wom_prob[1,8:23]

e0_wom_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0F.xlsx",sep="")),sheet = 4,startRow = 17)
e0_wom_prob_up_80 <- e0_wom_prob[1,8:23]

e0_wom_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0F.xlsx",sep="")),sheet = 1,startRow = 17)
e0_wom_prob_low_95 <- e0_wom_prob[1,8:23]

e0_wom_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_e0F.xlsx",sep="")),sheet = 5,startRow = 17)
e0_wom_prob_up_95 <- e0_wom_prob[1,8:23]

##
### 2.3.1 Total fertility rate
##

tfr_2 <- read.xlsx(file.path(the.data.path,
paste("WPP2019_FERT_F04_TOTAL_FERTILITY.xlsx",sep="")),
sheet = params$sheet_forecast,startRow = 17)
tfr_2_select <- tfr_2[1,8:23]
```

```
tfr_2_select_world_regions <- tfr_2[18:23,8:23]

##
### 2.3.2 Total fertility rate (median, 80% and 95% prediction intervals)
##

tfr_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_TFR.xlsx",sep="")),sheet = 3,startRow = 17)
tfr_prob_median <- tfr_prob[1,8:23]

tfr_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_TFR.xlsx",sep="")),sheet = 2,startRow = 17)
tfr_prob_low_80 <- tfr_prob[1,8:23]

tfr_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_TFR.xlsx",sep="")),sheet = 4,startRow = 17)
tfr_prob_up_80 <- tfr_prob[1,8:23]

tfr_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_TFR.xlsx",sep="")),sheet = 1,startRow = 17)
tfr_prob_low_95 <- tfr_prob[1,8:23]

tfr_prob <- read.xlsx(file.path(the.data.path,
paste("UN_PPP2019_Input_TFR.xlsx",sep="")),sheet = 5,startRow = 17)
tfr_prob_up_95 <- tfr_prob[1,8:23]
```

Brief data description. All data objects above contain forecasts from 2020 to 2100. All forecasts are available by region, some of them are available by single years of time, others by five years of time. Explore the data yourself.

3. Prepare population data

```
world_regions <- c("WORLD","Africa","Asia","Europe",
"Latin America and the Caribbean","Northern America","Oceania")
```

3.1 Create data object (array) that contains observed population data by world region, separately for women and men

Arrays can be a nice way to organize your data. Below you find an example to store observed population counts by age, calendar year, and world region:

```
row.names <- as.character(c(0:100))
column.names <- as.character(c(1950:2020))
matrix.names <- world_regions
wom_unwpp2019_1950_2020 <- array(NA,dim=c(101,71,7),
dimnames = list(row.names,column.names,matrix.names))
men_unwpp2019_1950_2020 <- array(NA,dim=c(101,71,7),
dimnames = list(row.names,column.names,matrix.names))

for(wr in 1:7){
    current.wr <- world_regions[wr]
    current.data.wom <- t(wom_select[which(wom_select[,1]==current.wr),])
```

```
    current.data.men <- t(men_select[which(men_select[,1]==current.wr),])
    wom_unwpp2019_1950_2020[,,wr] <- as.numeric(current.data.wom[3:103,])
    men_unwpp2019_1950_2020[,,wr] <- as.numeric(current.data.men[3:103,])
}
```

Please explain: *array()*, *for(){}*, *t()*, *which()*, and *as.numeric()*.

3.2 Create data object (array) that contains forecasted population data by world region, separately for women and men

And, of course, you can also store forecasted population data in such an array:

```
row.names <- as.character(c(0:100))
column.names <- as.character(c(2020:2100))
matrix.names <- world_regions
wom_unwpp2019_2020_2100 <- array(NA,dim=c(101,81,7),
dimnames = list(row.names,column.names,matrix.names))
men_unwpp2019_2020_2100 <- array(NA,dim=c(101,81,7),
dimnames = list(row.names,column.names,matrix.names))

for(wr in 1:7){
    current.wr <- world_regions[wr]
    current.data.wom <- t(wom_select_2[which(wom_select_2[,1]==current.wr),])
    current.data.men <- t(men_select_2[which(men_select_2[,1]==current.wr),])
    wom_unwpp2019_2020_2100[,,wr] <- as.numeric(current.data.wom[3:103,])
    men_unwpp2019_2020_2100[,,wr] <- as.numeric(current.data.men[3:103,])
}
```

How would an array look like that contains observed and forecasted population counts by age, calendar year, world region, and gender?

4. Analyze global population size

To have a closer look at the global population forecasts (deterministic and propbabilistic) we visualize them.

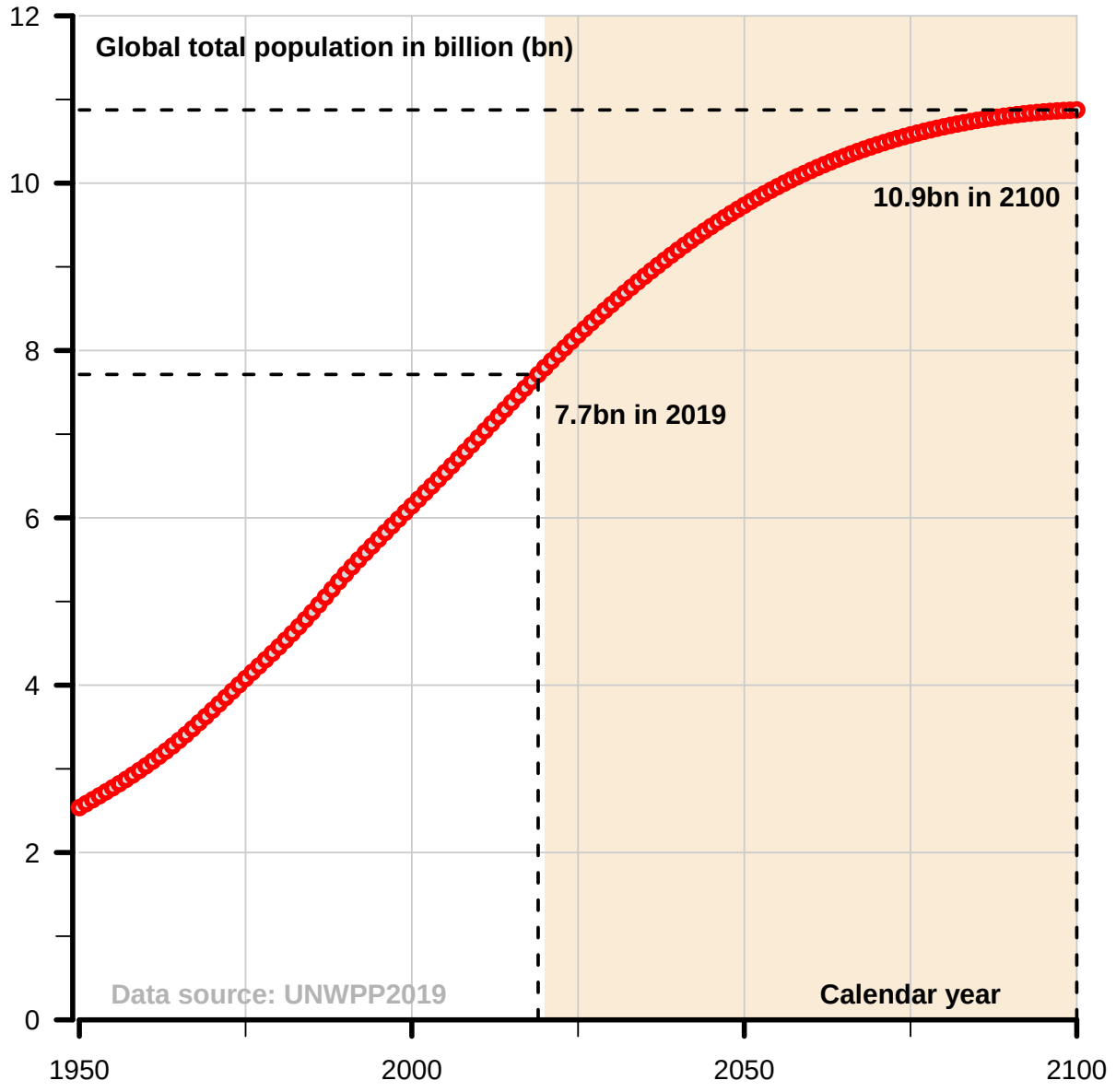
4.1 Display global population size, 1950 through 2100, based on deterministic forecast

```
setwd(the.plot.path)

dev.off()

pdf(file="global-world-pop-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```
require(wesanderson)
pal <- c("navy",wes_palette("Darjeeling1"))

par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(0,12),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=0, ytop=12, col="antiquewhite", border=NA)
```

```
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=0,y1=12,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(2,12,2),y1=seq(2,12,2),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=0)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=0)
    axis(side=2,at=seq(0,12,1),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(0,12,2),labels=TRUE,lwd=3,pos=1949)

    text(x=1950,y=11.6,"Global total population in billion (bn)",pos=4,font=2)
    text(x=2075,y=0,"Calendar year",pos=3,font=2)
    text(x=1980,y=0,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    ## World (UN release 1950:2020):

    lines(x=1950:2020,
y=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)[,"WORLD"]/1000000,
lwd=1,col=pal[2])

    points(x=1950:2020,
y=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)[,"WORLD"]/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    ## World (UN release 2020:2100):

    lines(x=2020:2100,
y=colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)[,"WORLD"]/1000000,
lwd=1,col=pal[2])

    points(x=2020:2100,
y=colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)[,"WORLD"]/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    segments(x0=2019,x1=2019,y0=0,
y1=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)["2019","WORLD"]/1000000,
col="black",lwd=2,lty=2)

    segments(x0=1950,x1=2019,
y0=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)["2019","WORLD"]/1000000,
y1=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)["2019","WORLD"]/1000000,
col="black",lwd=2,lty=2)

    text(x=2019,y=7.2,
paste(round(colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)["2019","WORLD"]/1000000,1),
"bn in 2019",sep=""),pos=4,font=2)

    segments(x0=2100,x1=2100,y0=0,
y1=colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)["2100","WORLD"]/1000000,
col="black",lwd=2,lty=2)

    segments(x0=1950,x1=2100,
y0=colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)["2100","WORLD"]/1000000,
y1=colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)["2100","WORLD"]/1000000,
col="black",lwd=2,lty=2)
```

```
    text(x=2100,y=9.8,
paste(round(colSums(wom_unwpp2019_2020_2100+men_unwpp2019_2020_2100)["2100","WORLD"]/1000000,1),
"bn in 2100",sep=""),pos=2,font=2)
```




```
dev.off()
```

Please explain: *par()*, *las=*, *mai=*, *plot()*, *axis()*, *seq()*, *rect()*, *segments()*, *pch=*, *lines()*, *points()*, *text()*, *pdf()*, *dev.off()*, *gray(0.85)*.

Brief data description. The red dotted line depicts the global population size (in billion) from 1950 to 2100.

Something to think about. 10.9bn people in 2100—do you remember how many people have been forecasted to live in 2100 in previous revisions of the UNWPP?

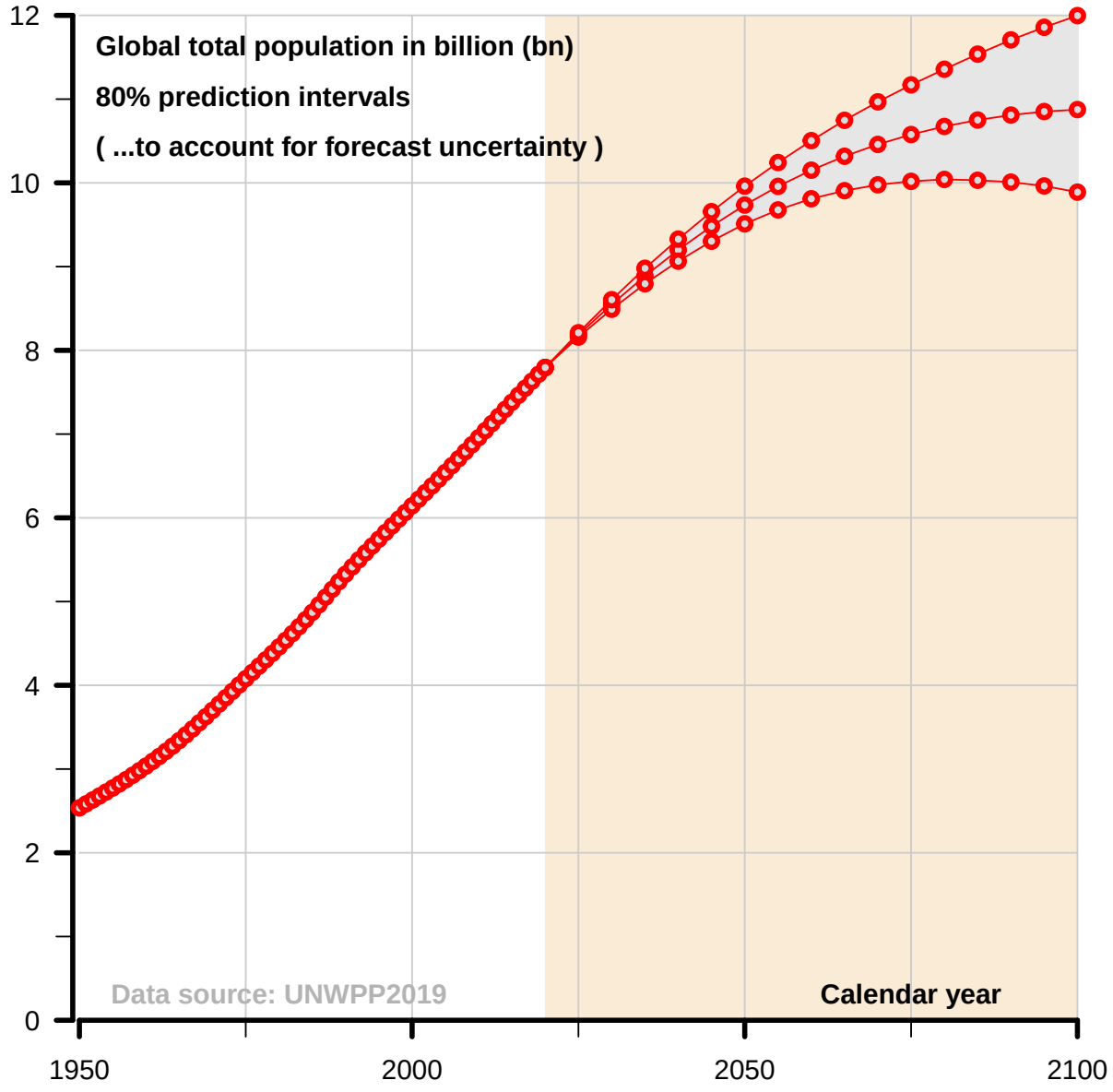
4.2 Display global population size, 1950 through 2100, based on probabilistic forecast

Probabilistic forecasts capture and quantify forecast uncertainty.

```
setwd(the.plot.path)

dev.off()

pdf(file="global-world-pop-prob-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```
require(wesanderson)
pal <- c("navy",wes_palette("Darjeeling1"))

par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(0,12),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=0, ytop=12, col="antiquewhite", border=NA)
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=0,y1=12,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(2,12,2),y1=seq(2,12,2),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=0)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=0)
    axis(side=2,at=seq(0,12,1),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(0,12,2),labels=TRUE,lwd=3,pos=1949)

    text(x=1950,y=11.6,"Global total population in billion (bn)",pos=4,font=2)
    text(x=1950,y=11.0,"80% prediction intervals",pos=4,font=2)
    text(x=1950,y=10.4,"( ...to account for forecast uncertainty )",pos=4,font=2)
    text(x=2075,y=0,"Calendar year",pos=3,font=2)
    text(x=1980,y=0,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    polygon(x=c(seq(2020,2100,5) , seq(2100,2020,-5)) ,
y=c(as.numeric(total_pop_low_80)/1000000 , rev(as.numeric(total_pop_up_80)/1000000)) ,
col=gray(0.9) , border=gray(0.9))

    ## World (UN release 1950:2020):
    lines(x=1950:2020,
y=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)[,"WORLD"]/1000000,
lwd=1,col=pal[2])

    points(x=1950:2020,
y=colSums(wom_unwpp2019_1950_2020+men_unwpp2019_1950_2020)[,"WORLD"]/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    ## World medium (UN release 2020:2100):
    lines(x=seq(2020,2100,5), y=as.numeric(total_pop_median)/1000000,lwd=1,col=pal[2])

    points(x=seq(2020,2100,5), y=as.numeric(total_pop_median)/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    ## World 80% pi (UN release 2020:2100):
    lines(x=seq(2020,2100,5), y=as.numeric(total_pop_low_80)/1000000,lwd=1,col=pal[2])
```

```
    points(x=seq(2020,2100,5), y=as.numeric(total_pop_low_80)/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(2020,2100,5), y=as.numeric(total_pop_up_80)/1000000,lwd=1,col=pal[2])

    points(x=seq(2020,2100,5), y=as.numeric(total_pop_up_80)/1000000,
pch=21,lwd=3,bg=gray(0.85),col=pal[2])
```




```
dev.off()
```

Please explain: *polygon()*.

Brief data description. The grey area depicts the 80% prediction interval of the probabilistic forecast. The line in the center represents the median, the upper and lower lines represent the upper and lower boundary of the 80% prediction interval.

Something to think about. How do you interpret the probabilistic forecast of the global population size in contrast to the deteministic forecast? What does the 80% prediction interval actually mean?

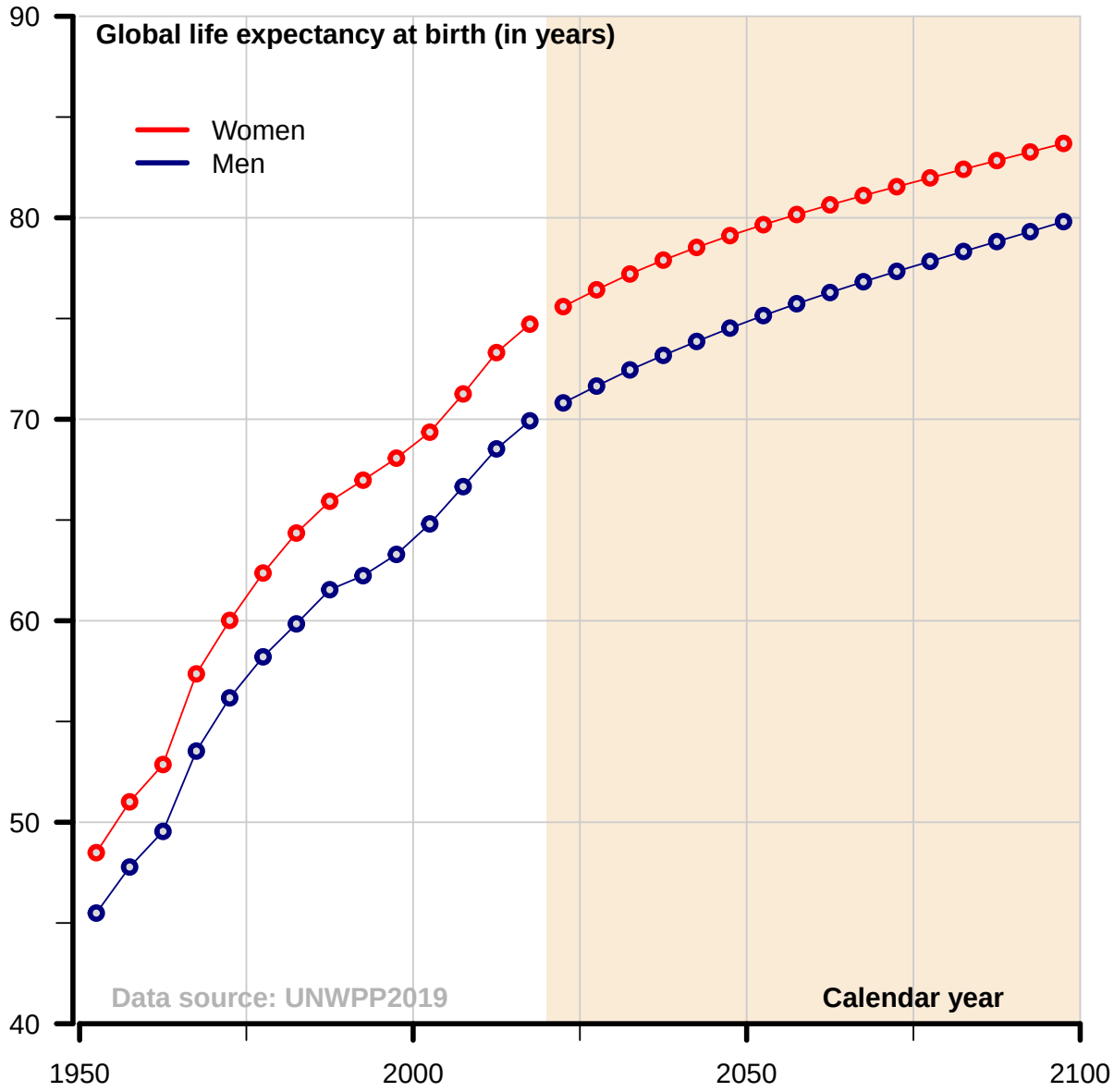
5. Analyze global life expectancy at birth

Following the procedure for the global population size above, we depict global life expectancy at birth for women and men based on the deterministic and probabilistic forecasts.

5.1 Display global life expectancy at birth for women and men based on deterministic forecast

```
setwd(the.plot.path)

dev.off()

pdf(file="global-world-lexp-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```
par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(40,90),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=40, ytop=90, col="antiquewhite", border=NA)
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=40,y1=90,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(40,90,10),y1=seq(40,90,10),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=40)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=40)
    axis(side=2,at=seq(40,90,5),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(40,90,10),labels=TRUE,lwd=3,pos=1949)

    text(x=1950,y=89,"Global life expectancy at birth (in years)",pos=4,font=2)
    text(x=2075,y=40,"Calendar year",pos=3,font=2)
    text(x=1980,y=40,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    ## World (UN release 1950:2020):
    lines(x=seq(1952.5,2017.5,5), y=e0_wom_select,lwd=1,col=pal[2])
    points(x=seq(1952.5,2017.5,5), y=e0_wom_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(1952.5,2017.5,5), y=e0_men_select,lwd=1,col=pal[1])
    points(x=seq(1952.5,2017.5,5), y=e0_men_select,pch=21,lwd=3,bg=gray(0.85),col=pal[1])

    ## World (UN release 2020:2100):
    lines(x=seq(2022.5,2097.5,5), y=e0_wom_2_select,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=e0_wom_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=e0_men_2_select,lwd=1,col=pal[1])
    points(x=seq(2022.5,2097.5,5), y=e0_men_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[1])

    legend(x=1955,y=86,c("Women","Men"),col=c(pal[2],pal[1]),lty=1,bty="n",lwd=3)
```




```
dev.off()
```

Please explain: *legend()*.

Brief data description. The red and blue dotted lines depict global life expectancy at birth (in years) for women and men, respectively, from 1950 to 2100.

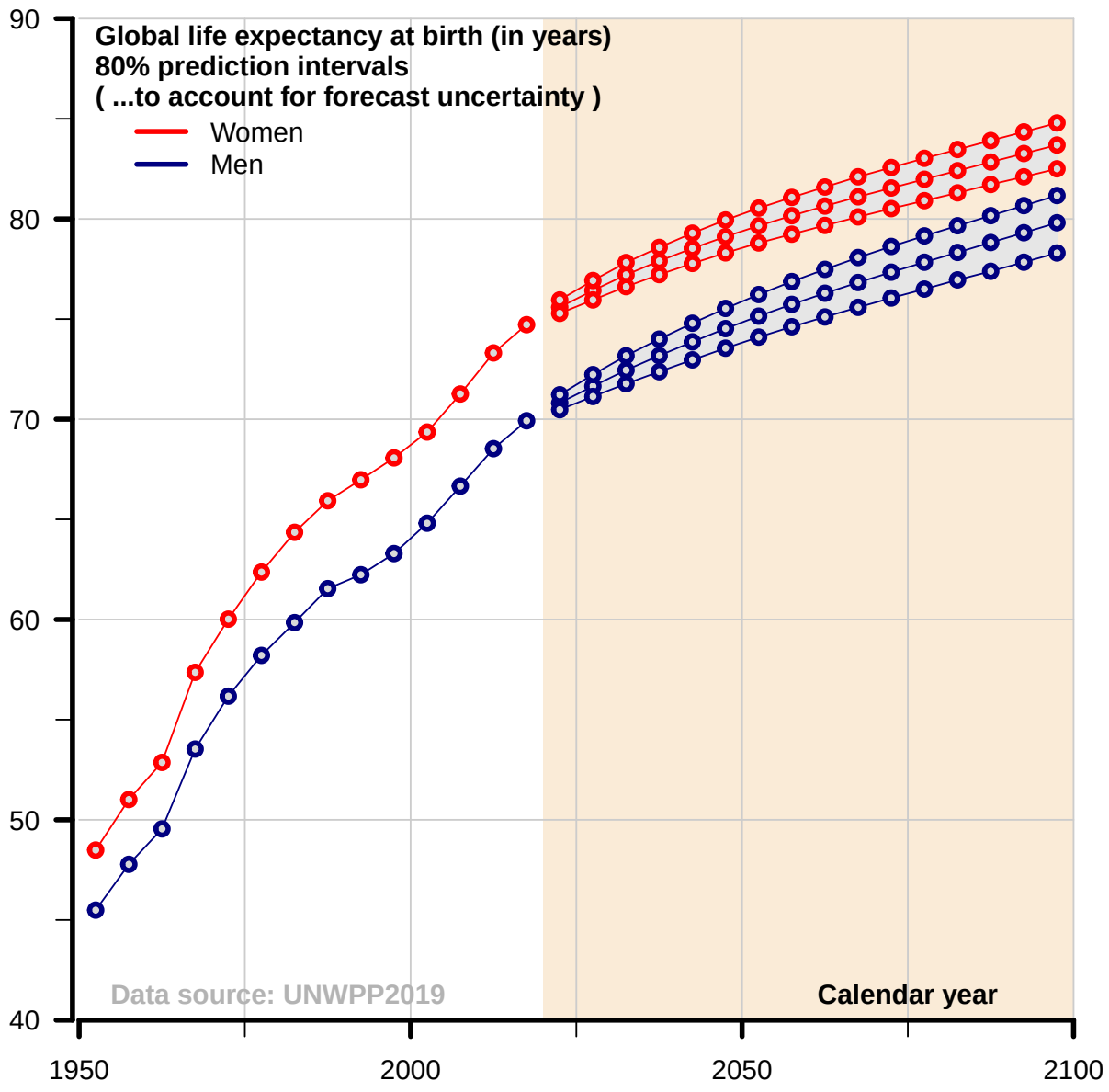
5.2 Display global life expectancy at birth for women and men based on probabilistic forecast

```
setwd(the.plot.path)
```

```
dev.off()

pdf(file="global-world-lexp-prob-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```
par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(40,90),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=40, ytop=90, col="antiquewhite", border=NA)
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=40,y1=90,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(40,90,10),y1=seq(40,90,10),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=40)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=40)
    axis(side=2,at=seq(40,90,5),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(40,90,10),labels=TRUE,lwd=3,pos=1949)

    text(x=1950,y=89,"Global life expectancy at birth (in years)",pos=4,font=2)
    text(x=1950,y=87.5,"80% prediction intervals",pos=4,font=2)
    text(x=1950,y=86,"( ...to account for forecast uncertainty )",pos=4,font=2)
    text(x=2075,y=40,"Calendar year",pos=3,font=2)
    text(x=1980,y=40,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    polygon(x=c(seq(2022.5,2097.5,5) , seq(2097.5,2022.5,-5)) ,
y=c(e0_wom_prob_low_80 , rev(e0_wom_prob_up_80 ) ) , col=gray(0.9) , border=gray(0.9))

    polygon(x=c(seq(2022.5,2097.5,5) , seq(2097.5,2022.5,-5)) ,
y=c(e0_men_prob_low_80 , rev(e0_men_prob_up_80 ) ) , col=gray(0.9) , border=gray(0.9))

    ## World (UN release 1950:2020):
    lines(x=seq(1952.5,2017.5,5), y=e0_wom_select,lwd=1,col=pal[2])
    points(x=seq(1952.5,2017.5,5), y=e0_wom_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(1952.5,2017.5,5), y=e0_men_select,lwd=1,col=pal[1])
    points(x=seq(1952.5,2017.5,5), y=e0_men_select,pch=21,lwd=3,bg=gray(0.85),col=pal[1])

    ## World (UN release 2020:2100):
    lines(x=seq(2022.5,2097.5,5), y=e0_wom_2_select,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=e0_wom_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=e0_wom_prob_low_80 ,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=e0_wom_prob_low_80 ,pch=21,lwd=3,
bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=e0_wom_prob_up_80,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=e0_wom_prob_up_80,pch=21,lwd=3,
bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=e0_men_2_select,lwd=1,col=pal[1])
    points(x=seq(2022.5,2097.5,5), y=e0_men_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[1])

    lines(x=seq(2022.5,2097.5,5), y=e0_men_prob_low_80 ,lwd=1,col=pal[1])
    points(x=seq(2022.5,2097.5,5), y=e0_men_prob_low_80 ,pch=21,lwd=3,
```

```
bg=gray(0.85),col=pal[1])

    lines(x=seq(2022.5,2097.5,5), y=e0_men_prob_up_80,lwd=1,col=pal[1])
    points(x=seq(2022.5,2097.5,5), y=e0_men_prob_up_80,pch=21,lwd=3,
bg=gray(0.85),col=pal[1])

    legend(x=1955,y=86,c("Women","Men"),col=c(pal[2],pal[1]),lty=1,bty="n",lwd=3)
```




```
dev.off()
```

Brief data description. The grey area depicts the 80% prediction interval of the probabilistic forecast. The line in the center represents the median, the upper and lower lines represent the upper and lower boundary of the 80% prediction interval.

Something to think about. How do you interpret the probabilistic forecast of global life expectancy at birth in contrast to the deteministic forecast? What does the 80% prediction interval actually mean?

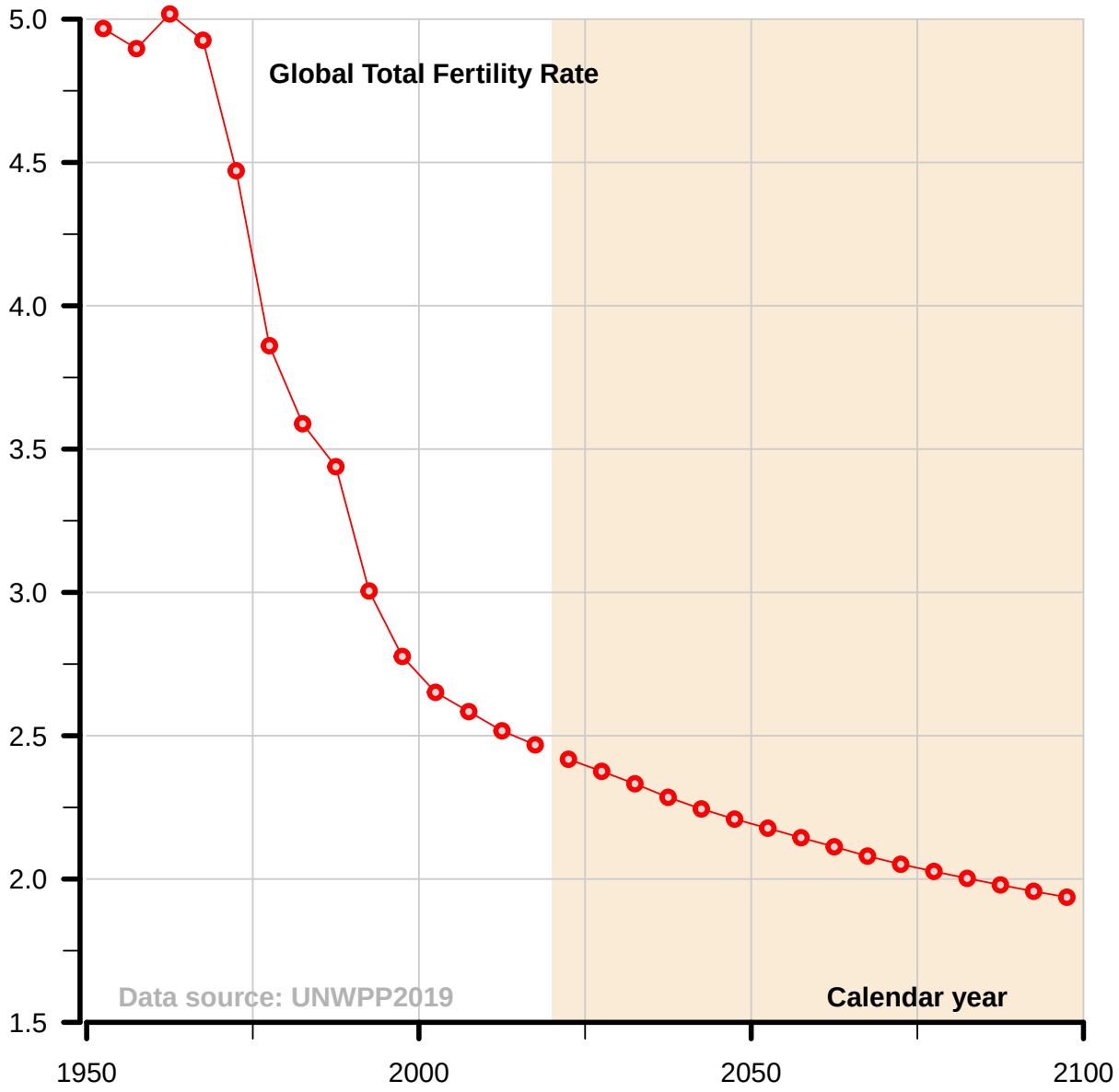
6. Analyze global total fertility rate

Following the procedure for the global population size and life expectancy at birth above, we depict the global total fertility rate (TFR) for women and men based on the deterministic and probabilistic forecasts.

6.1 Display global total fertility rate based on deterministic forecast

```
setwd(the.plot.path)

dev.off()

pdf(file="global-world-tfr-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```
par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(1.5,5),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=1.5, ytop=5, col="antiquewhite", border=NA)
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=1.5,y1=5,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(1.5,5,0.5),y1=seq(1.5,5,0.5),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=1.5)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=1.5)
    axis(side=2,at=seq(1.5,5,0.25),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(1.5,5,0.5),labels=TRUE,lwd=3,pos=1949)

    text(x=1950+25,y=4.8,"Global Total Fertility Rate",pos=4,font=2)
    text(x=2075,y=1.5,"Calendar year",pos=3,font=2)
    text(x=1980,y=1.5,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    ## World (UN release 1950:2020):
    lines(x=seq(1952.5,2017.5,5), y=tfr_select,lwd=1,col=pal[2])
    points(x=seq(1952.5,2017.5,5), y=tfr_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    ## World (UN release 2020:2100):
    lines(x=seq(2022.5,2097.5,5), y=tfr_2_select,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=tfr_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])
```



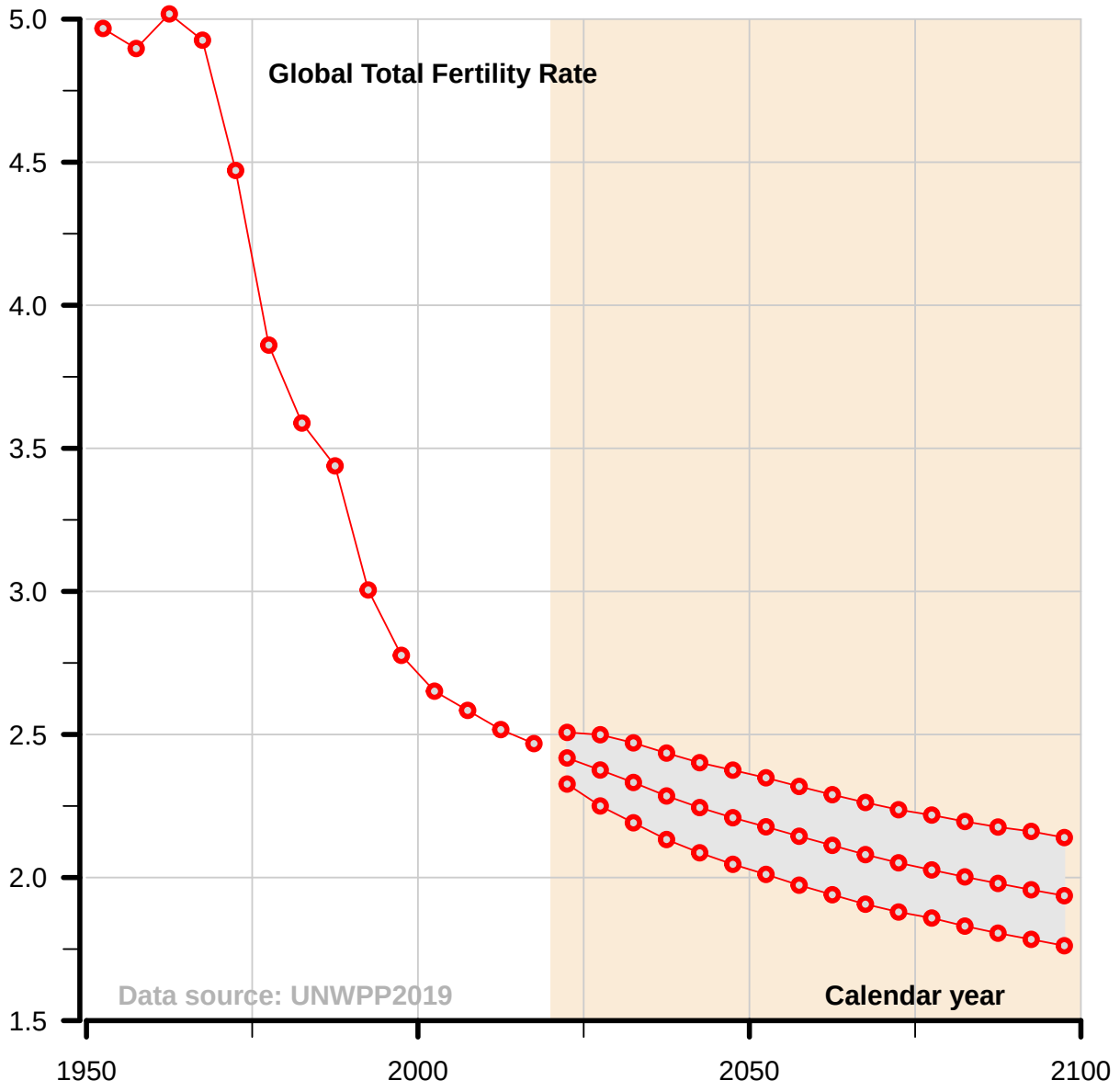

```
dev.off()
```

Brief data description. The red dotted lines represents the global total fertility rate from 1950 to 2100.

6.2 Display global total fertility rate based on probabilistic forecast

```
setwd(the.plot.path)

dev.off()

pdf(file="global-world-tfr-prob-1950-2100.pdf", width=10, height=10, family="Times",
pointsize=20, onefile=TRUE)
```

```r
par(fig = c(0,1,0,1), las=1, mai=c(0.4,0.4,0.0,0))

    plot(x=-100,y=-100,xlim=c(1949,2100),ylim=c(1.5,5),xlab="",ylab="",main="",axes=FALSE)

    rect(xleft=2020, xright=2100, ybottom=1.5, ytop=5, col="antiquewhite", border=NA)
    segments(x0=seq(1975,2100,25),x1=seq(1975,2100,25),y0=1.5,y1=5,col=gray(0.8))
    segments(x0=1950,x1=2100,y0=seq(1.5,5,0.5),y1=seq(1.5,5,0.5),col=gray(0.8))

    axis(side=1,at=seq(1950,2100,25),labels=FALSE,lwd=1,pos=1.5)
    axis(side=1,at=seq(1950,2100,50),labels=TRUE,lwd=3,pos=1.5)
    axis(side=2,at=seq(1.5,5,0.25),labels=FALSE,lwd=1,pos=1949)
    axis(side=2,at=seq(1.5,5,0.5),labels=TRUE,lwd=3,pos=1949)

    text(x=1950+25,y=4.8,"Global Total Fertility Rate",pos=4,font=2)
    text(x=2075,y=1.5,"Calendar year",pos=3,font=2)
    text(x=1980,y=1.5,"Data source: UNWPP2019",pos=3,font=2,col=gray(0.7))

    polygon(x=c(seq(2022.5,2097.5,5) , seq(2097.5,2022.5,-5)) ,
y=c(tfr_prob_low_80 , rev(tfr_prob_up_80 ) ) , col=gray(0.9) , border=gray(0.9))

    ## World (UN release 1950:2020):
    lines(x=seq(1952.5,2017.5,5), y=tfr_select,lwd=1,col=pal[2])
    points(x=seq(1952.5,2017.5,5), y=tfr_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    ## World (UN release 2020:2100):
    lines(x=seq(2022.5,2097.5,5), y=tfr_2_select,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=tfr_2_select,pch=21,lwd=3,bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=tfr_prob_low_80,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=tfr_prob_low_80,pch=21,lwd=3,
bg=gray(0.85),col=pal[2])

    lines(x=seq(2022.5,2097.5,5), y=tfr_prob_up_80,lwd=1,col=pal[2])
    points(x=seq(2022.5,2097.5,5), y=tfr_prob_up_80,pch=21,lwd=3,
bg=gray(0.85),col=pal[2])
```




```
dev.off()
```

Brief data description. The grey area depicts the 80% prediction interval of the probabilistic forecast. The line in the center represents the median, the upper and lower lines represent the upper and lower boundary of the 80% prediction interval.

Something to think about. How do you interpret the probabilistic forecast of the global total fertility rate in contrast to the deteministic forecast? What does the 80% prediction interval actually mean?

7. Now is the time to do this analysis again (sections 4 through 6) for your assigned world region. Feel free to change and adapt R code to your needs and creativity :-)