Assignment3

Miao Li (18230232)/Jing Loiu(18231917)

2019.3.3.

Introduction:

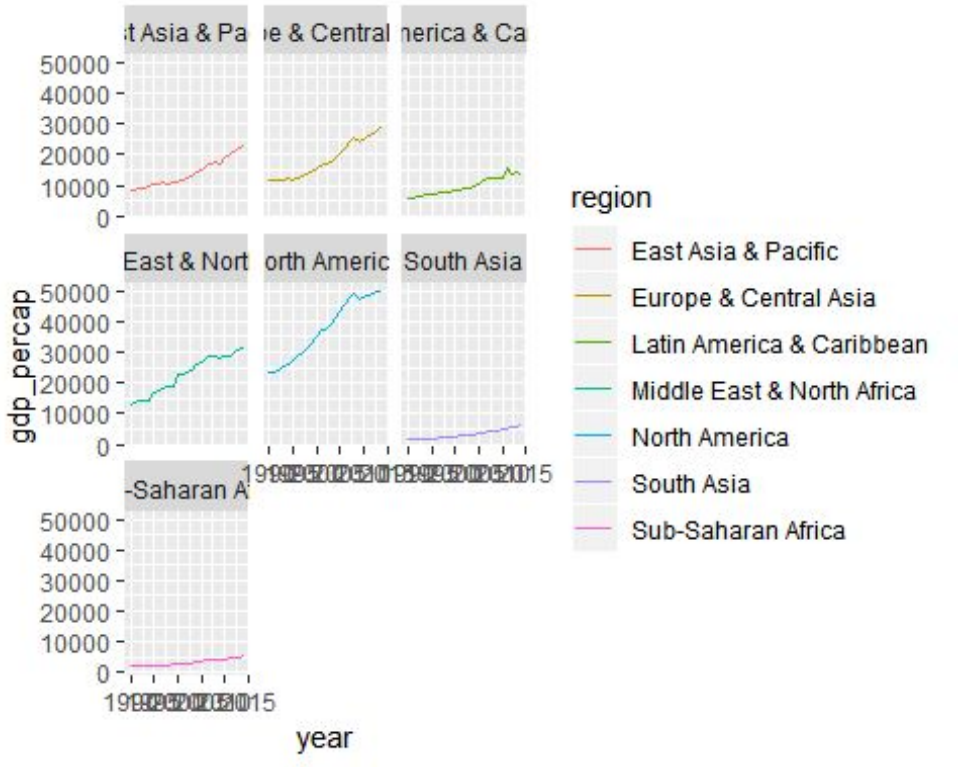
There are two parts in this assignment, the first part is to detect the connection among life expectancy, regions and GDP; The second part is to prove if Ridley's theory is right; Contribution: Jing Liu: question a,b,5; Miao Li 3,4;

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.2

library(lattice)
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

library(ggplot2)
library(dplyr)

worldpopulation<-read.csv("worldpopulation_percentage.csv")
nations<-read.csv("nations.csv")
children<-read.csv("children-per-woman-UN.csv")

nations<-nations[nations$income!='Not classified',]
```

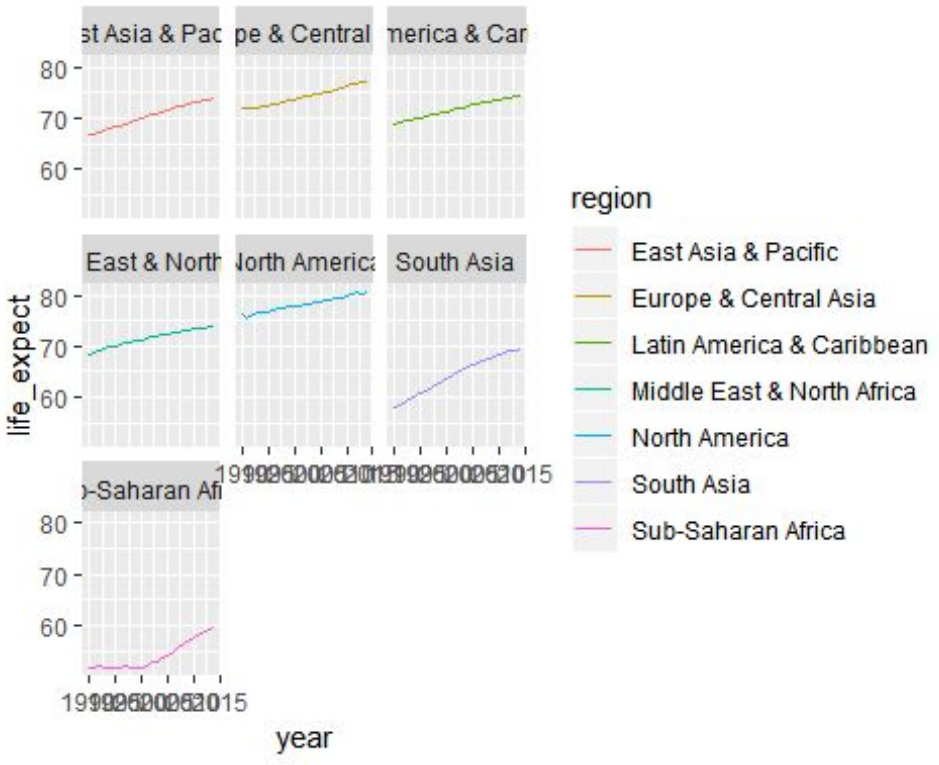
1. there are totally 9 grids for 9 different regions, from the plot we can see all region's gdp are improved with the increasing of years, and same to life expectation, the life expectancyof each region improved with the increase of years.for the average GDP an average life expectancycomparison on plot3, each region can leaves a trial on the plot,each region when the GDP increasing the life expectancyincreasing as well, so the goal of each region especially the north American enjoys a very higher life expectancyand GDP.The size of the points indicate very well that a small proportion of the world's population enjoys the twin benefits of high GDP and high life expectancy, such as American, but some of region such as sub-saharan Afica haven't achieved high life expectancyand

GDP but it is rapidly accelerating in this direction.for some region who has big population like east Asian and south Asian and Latin American on the overlapping area on plot3, most of countries are developing country in these regions, even the life expectancyand GDP are not higher than developed region like Europ and north American, but they also rapidly accelerating in this direction.

```
nation<-read.csv('nations.csv')
nation1<-nations%>%select(year,region,country,gdp_percap)%>%filter(!is.
na(gdp_percap))%>%group_by(region,year)%>%summarise(gdp_percap=mean(gdp
_percap))
ggplot(nation1,aes(x=year,y=gdp_percap,color=region))+geom_line()+facet
_wrap(~region)
```

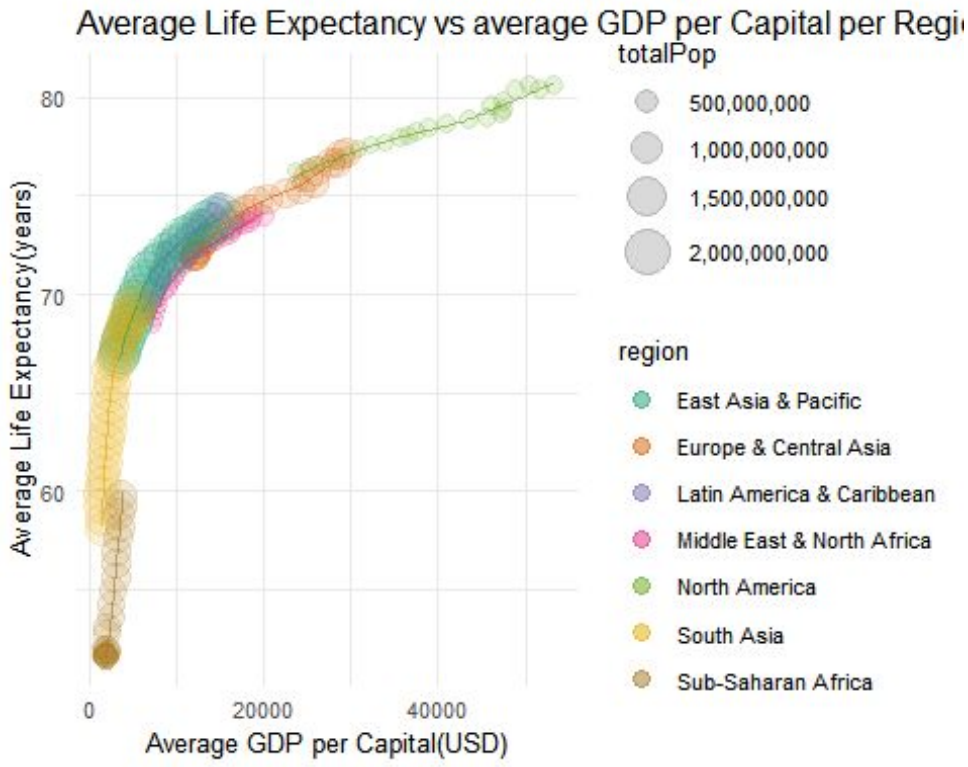



```
#ggplot(nation1,aes(x=year,y=gdp_percap,color=region))+geom_line()

nation2<-nations%>%select(year,region,country,life_expect)%>%filter(!is.
na(life_expect))%>%group_by(region,year)%>%summarise(life_expect=mean(l
ife_expect))
ggplot(nation2,aes(x=year,y=life_expect,color=region))+geom_line()+face
t_wrap(~region)
```




```
#ggplot(nation2,aes(x=year,y=life_expect,color=region))+geom_line()

library(ggrepel)

## Warning: package 'ggrepel' was built under R version 3.5.2

library(scales)
regions<-nations%>% select(region,year,population,gdp_percap,life_expec
t)%>%group_by(year,region)%>%summarise(totalPop=sum(population,na.rm=TR
UE), ave_gdp_percap=weighted.mean(gdp_percap,population,na.rm=TRUE),ave
_life_expect=weighted.mean(life_expect,na.rm=TRUE))

ggplot(regions,aes(x=ave_gdp_percap,y=ave_life_expect,size=totalPop,col
or=region))+
    geom_point(na.rm = TRUE,alpha=0.15)+
    ggtitle('Average Life Expectancy vs average GDP per Capital per Regi
on ')+
    xlab('Average GDP per Capital(USD)')+
    ylab('Average Life Expectancy(years)')+
    scale_color_brewer(palette='Dark2')+
    scale_size_area(max_size = 8, labels  = comma)+
    theme_minimal(base_size=10) +
    # increase the size of the legend colour points, and their alpha val
ue (too fain otherwise)
    guides(color = guide_legend(override.aes = list(size = 3, alpha = 0.
5))) +
```

```
    geom_line(size = 0.5  , stat="smooth",method = "loess",se=FALSE, sho
w.legend =FALSE, na.rm = TRUE,  alpha = 0.6)
```



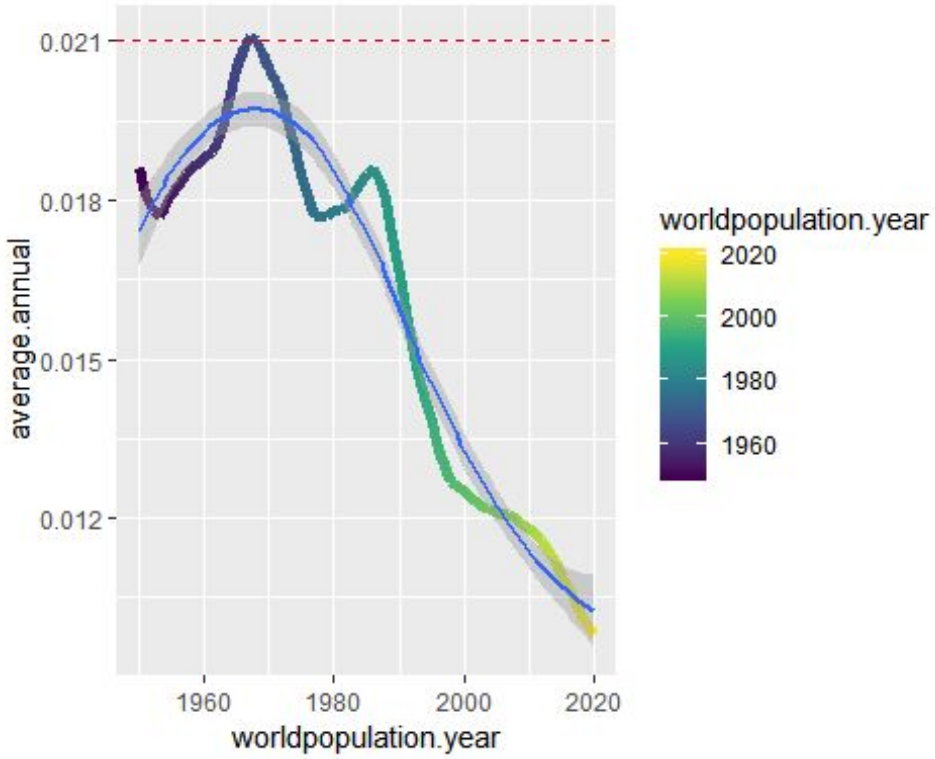
3. Percentage increase in world population

a. Use “worldpopulation_percentage” dataset to get the data of year and percentage increase; This dataset has four columns “year”,“population”, “average annual” and “average annual”. We chose to show “year” and “average annual” to see the trend of population growth in percentage.

Result: the result seems prove Ridley’s theory. As you can see that birth rate keeps decreasing after 1965, although it has fluctuations during that period. In 2000, the birth rate is only 0.014. After 2008, birth rate is under 0.0105. Also this graph makes prediction that the birth rate will continue decrease after 2019. Overall, birth rate is decreasing, and it is lower then 0.012 in recent years. Therefore, personally, I think Ridley’s theory is right that the the population will shrink over time.

```
#use Iapply to convert % in to numeric
value <- data.frame(lapply(worldpopulation["average.annual"], function
(x) as.numeric(sub("%", "", x))/100) )
world_population_growth <- data.frame(worldpopulation$year,value)
ggplot(world_population_growth, aes(worldpopulation.year, average.annua
l,colour= worldpopulation.year)) +
  scale_colour_viridis_c() +
  geom_line(aes(colour = worldpopulation.year),size = 2)+geom_smooth()+
geom_hline(aes(yintercept=0.021),col="red",linetype="dashed")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



4. Changing fertility rates from 1950 to the present for selected countries of the world

a. Use "nations" dataset, chose "country", "year",

```
GDP_table <- data.frame(nations$country,nations$year,nations$gdp_percap,
nations$income,nations$region)

group_table<- GDP_table %>% group_by(GDP_table$nations.country,nations$
income,nations$region)%>%
  summarise(vag_GDP=mean(nations.gdp_percap,na.rm=T))
names(group_table) <- c("Entity","income","region","vag_GDP")
```

b. Use "mutate" and "case_when" to creat new column with four different types("Rich","Poor","Developing","Developed")

```
table <- nations%>%mutate(Type = case_when(
    nations$income== "High income"~ "Rich",
    nations$income== "Low income"~ "Poor",
    nations$income== "Upper middle income" ~ "Developing",
    nations$income== "Lower middle income"~ "Developed"
    ))
table2<- table %>% group_by(table$country,table$region,table$Type)%>%
  summarise(vag_GDP=mean(table$gdp_percap,na.rm=T))
names(table2) <- c("Entity","region","Type","vag_GDP")

zz <- inner_join(group_table,table2,by ="Entity" )
```

```
table3 <-subset(zz, select=c(Entity,region.x,vag_GDP.x,Type))
names(table3) <- c("Entity","region","vag_GDP","Type")
```

C. Select 10 countries for each type. For exmple, 10 poor countries, 10 rich countries.....

```
# 1. Use "subset" funtion to get four types ("poor","Rich","Developed","
Developing")
# 2. Use "sample" function to select 10 rows randomly
# 3. Use "is.na()" function to filt out "NaN"
# 4. Use "mutate_at" and ".vars" to change the type of columns
Poor <- subset(table3,table3$Type=="Poor"& !is.na(table3$vag_GDP)==T)
Poor_random<-Poor[sample(nrow(Poor),10,replace=F),]
Rich <- subset(table3,table3$Type=="Rich" & !is.na(table3$vag_GDP)==T)
Rich_random<-Rich[sample(nrow(Rich),10,replace=F),]
Developed <- subset(table3,table3$Type=="Developed" & !is.na(table3$vag
_GDP)==T)
Developed_random<-Developed[sample(nrow(Developed),10,replace=F),]
Developing <- subset(table3,table3$Type=="Developing" & !is.na(table3$v
ag_GDP)==T)
Developing_random<-Developing[sample(nrow(Developing),10,replace=F),]
GDP_four_countries <- rbind(Poor_random,Rich_random,Developed_random,De
veloping_random)
names(GDP_four_countries) <- c("Entity","region","vag_GDP","Type")
GDP_four_countries <- GDP_four_countries %>%
  mutate_at(.vars = vars(Entity,region,vag_GDP,Type), .fun = as.characte
r)
```

d. Get the final table which has six columns("Year","birth_year","Entity","Estimates","Type","region")

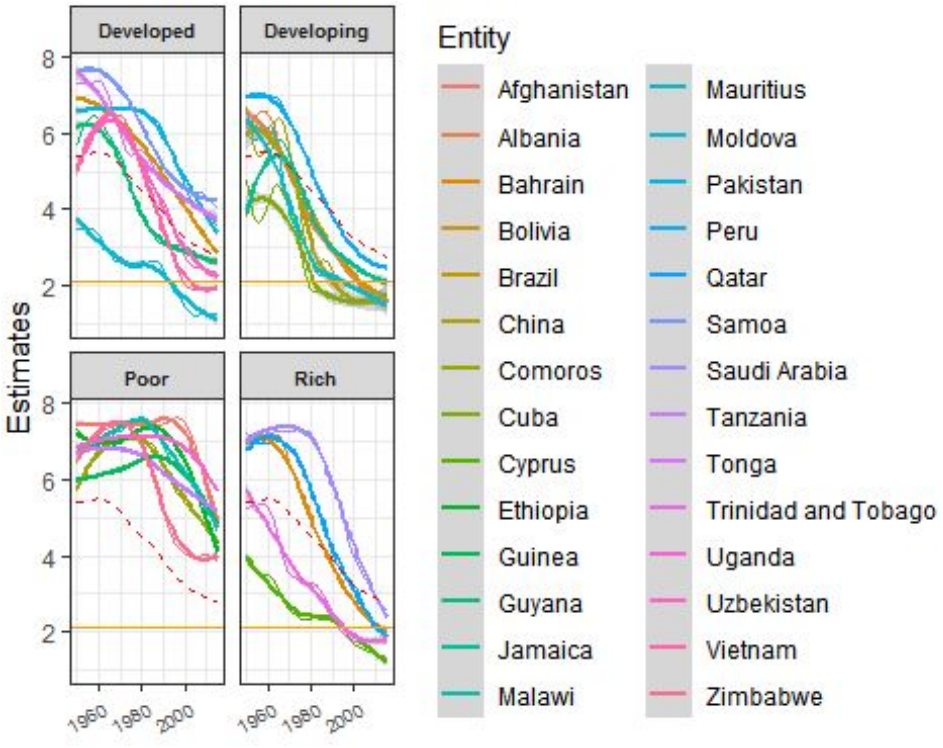
```
new_children <- children %>%
  mutate_at(.vars = vars(Entity), .fun = as.character) %>%
  mutate_at(.vars = vars(Estimates,Year), .fun = as.numeric)
new_children <- as_tibble(new_children)%>% select(Entity,Year,Estimates)
#the world fertility average (calculated yearly)
yearly_birth <- children%>% group_by(Year) %>%summarise(birth_year = me
an(Estimates))%>%select(Year,birth_year)
final_children <- inner_join(GDP_four_countries,new_children,by="Entity
")%>%select(Entity,Year,Estimates,Type,region)
final_children1 <- inner_join(yearly_birth,final_children,by="Year")
```

The result shows that the birth rate of four types of countries all decrease after 1980. All lines follow the red dashed line–"the world fertility average". In terms of poor countries, the birth rate are all higher then "the world fertility average". There are 2.1 orange lines in each graph, and it is clear that the birth rate of developing and rich is below 2.1 in recent years. But the birth rate of developed and poor countries is still above 2.1. In order to see the trend more clearly, we plot another graph. It shows that except poor countries, the birth rate of other three types of countries are

all around “the world fertility average” line. Only poor country has a higher birth rate then “the world fertility average” ;

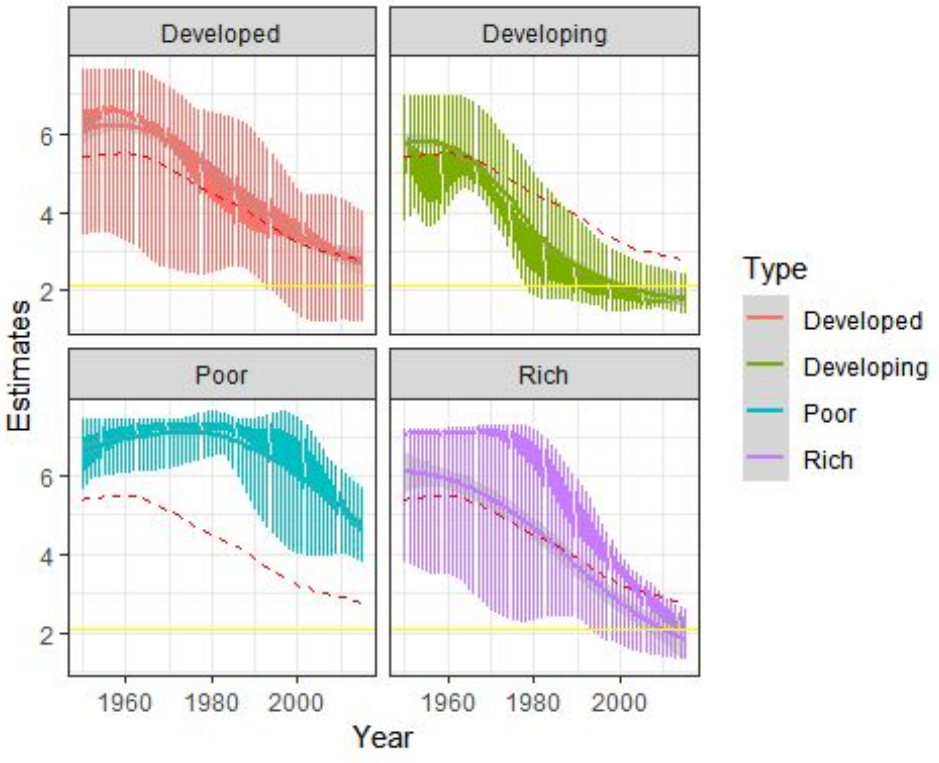
```
  ggplot(final_children1,aes(x =Year,y=Estimates, colour=Entity))+geom_
line()+theme_bw() +facet_wrap(~Type)+geom_smooth()+geom_line(data =fina
l_children1, aes(x =Year, y=birth_year), col="red", size = 0.7, linetype
="dashed" ) + geom_hline(aes(yintercept=2.1),col="orange")+theme(
        plot.margin = margin(14, 7, 3, 1.5),
        axis.text.x = element_text(angle=30, hjust = 1, size=7),
        axis.title.x = element_blank(),
        strip.text.x = element_text(size=7, face="bold"))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



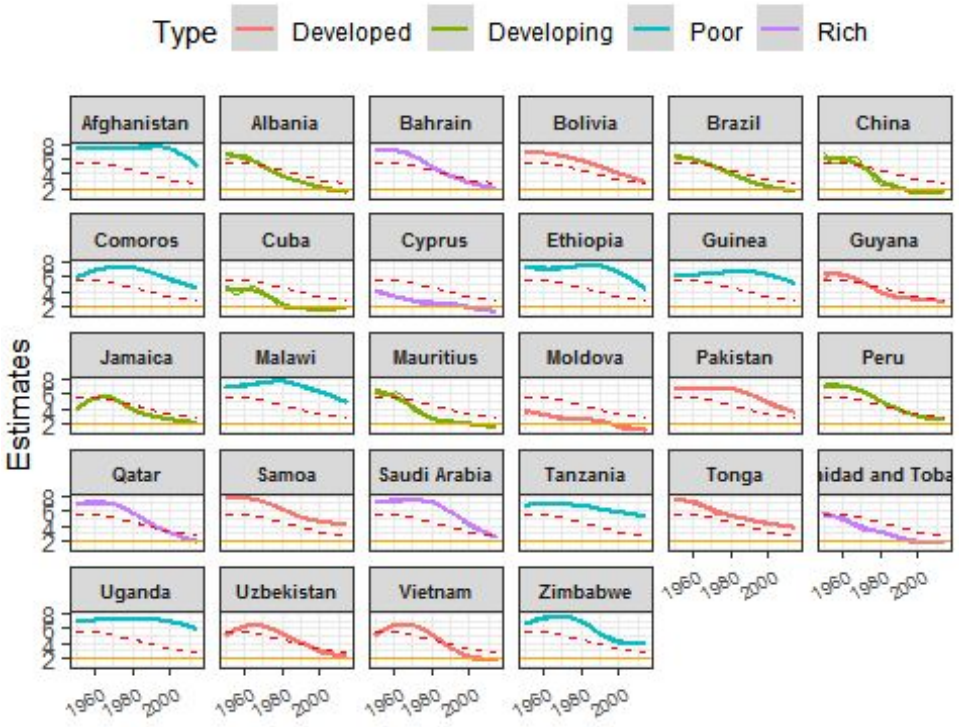

```
ggplot(final_children1,aes(x =Year,y=Estimates, colour=Type))+geom_line
()+theme_bw() +facet_wrap(~Type)+geom_smooth()+geom_line(data =final_ch
ildren1, aes(x =Year, y=birth_year), col="red", size = 0.7, linetype="da
shed" ) + geom_hline(aes(yintercept=2.1),col="yellow")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



More evidence: In this graph, it is clear that most countries are close to the "the world fertility average" line . Furthermore, line graph of every country decreases and most of them are close to 2.1 in recent years. It means that the birth rate of all over the world is decreasing and the birth rate of most countries will be lower then 2.1. Overall, I have enough confidence to say that Ridley's theory is right.

```
 ggplot(final_children1,aes(x =Year,y=Estimates, colour=Type))+geom_lin
e()+theme_bw() +facet_wrap(~Entity)+geom_smooth()+geom_line(data =final
_children1, aes(x =Year, y=birth_year), col="red", size = 0.7, linetype=
"dashed" ) + geom_hline(aes(yintercept=2.1),col="orange")+theme(legend.
position = "top",
       plot.margin = margin(14, 7, 3, 1.5),
       axis.text.x = element_text(angle=30, hjust = 1, size=7),
       axis.title.x = element_blank(),
       strip.text.x = element_text(size=7, face="bold"))

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



5. Use region to to plot graph.

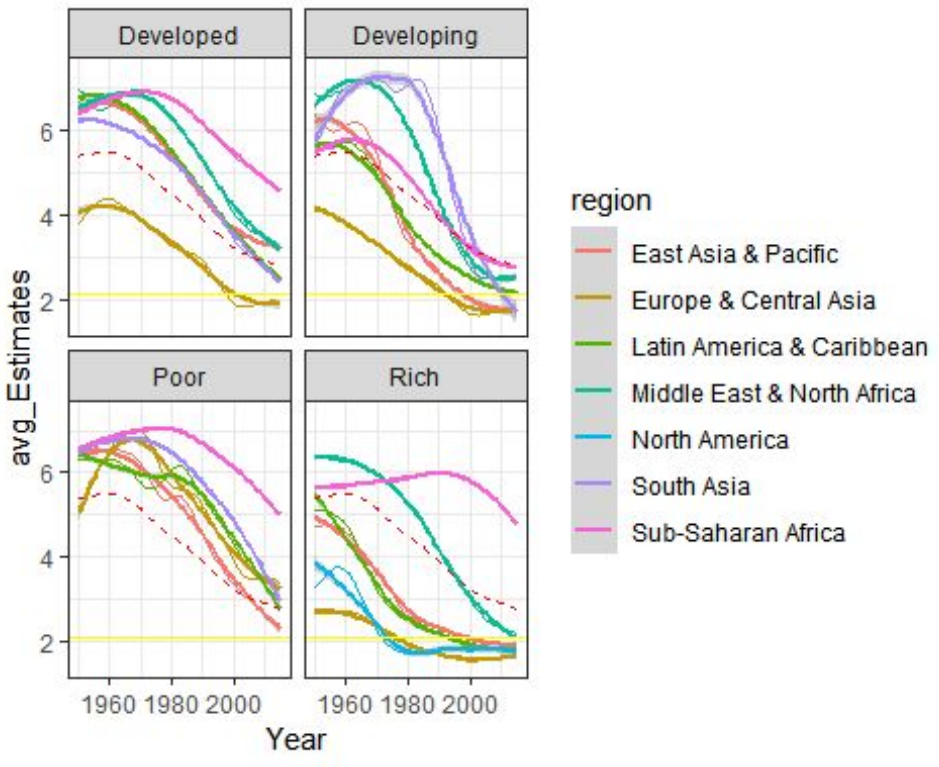
a. Group table by "region"

```
GDP_table_reagion2<- data.frame(nations$country,nations$year,nations$gd
p_percap,nations$income,nations$region)
reagion_group2<- GDP_table_reagion2 %>% group_by(nations$region)%>%
  summarise()
names(reagion_group2) <- c("region")

reagion_table <- left_join(reagion_group2,nations,by="region")%>%select
(region,country,year,income)
names(reagion_table) <- c("region","Entity","Year","income")

reagion_table$Entity<- as.character(reagion_table$Entity)
zzx<-reagion_table
tt1 <- reagion_table%>%group_by(Entity)%>%summarise()

tt2 <- left_join(tt1,new_children,by="Entity")
final_reagion_table_n <- inner_join(tt2,zzx,by="Entity")%>%select(Entit
y,Year.x,Estimates,region,income)
names(final_reagion_table_n) <- c("Entity","Year","Estimates","region",
"income")

final_reagion_table_n$Year<-as.integer(final_reagion_table_n$Year)
yearly_birth$Year<-as.integer(yearly_birth$Year)

#final_reagion_table_n%>%mutate_at(.vars = vars(Year), .fun = as.numeri
```

```
c)
#yearly_birth%>%mutate_at(.vars = vars(Year), .fun = as.numeric)
ttttt <- inner_join(final_reagion_table_n,yearly_birth,by="Year")
final_reagion_table_n1 <- ttttt%>%mutate(Type = case_when(
    ttttt$income== "High income"~ "Rich",
    ttttt$income== "High income: OECD"~ "Rich",
    ttttt$income== "Low income"~ "Poor",
    ttttt$income== "Upper middle income" ~ "Developing",
    ttttt$income== "Lower middle income"~ "Developed"

    ))

final_reagion_table_n2 <- final_reagion_table_n1 %>%group_by(region,Yea
r,Type)%>%summarise(avg_Estimates= mean(Estimates),avg_birth=mean(birth
_year))
```

The result shows that the decreasing trend can be seen in four types of country; In recent years, birth rate of developing and rich is below 2.1 line and “the world fertility average” line. Overall, birth rate keeps decreasing.More details can be shown in the next graph.

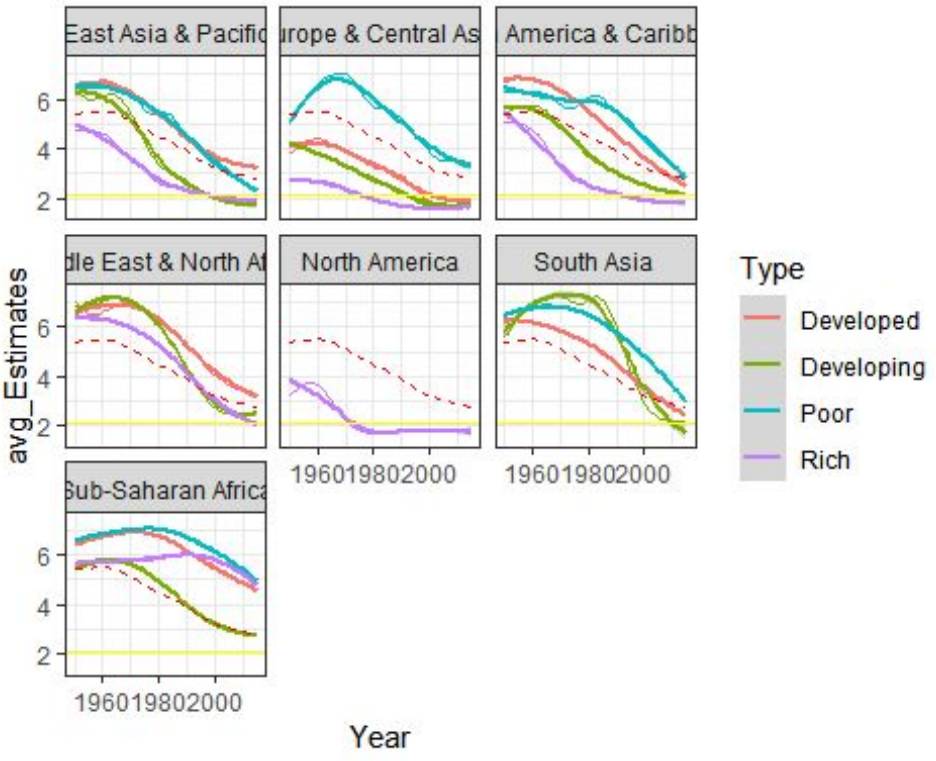
```
ggplot(final_reagion_table_n2,aes(x =Year,y=avg_Estimates,colour=regio
n))+geom_line()+theme_bw() +facet_wrap(~Type)+geom_smooth()+geom_line(d
ata =final_reagion_table_n2, aes(x =Year, y=avg_birth), col="red", size
= 0.7, linetype="dashed" ) + geom_hline(aes(yintercept=2.1),col="yellow
")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



It is clear that the birth rate of all regions decrease all the time. The most obvious trend is shown in developing and rich regions, where birth rate is close or below 2.1 in 2000. Based on what we plot, we can say that Ridley's theory is right. If there is no accident, fertility rates everywhere will converge to 2.1 in a few decades, and the world population will stabilise at 9 billion people.

```
ggplot(final_reagion_table_n2,aes(x =Year,y=avg_Estimates,colour=Type))
+geom_line()+theme_bw() +facet_wrap(~ region)+geom_smooth()+geom_line(d
ata =final_reagion_table_n2, aes(x =Year, y=avg_birth), col="red", size
= 0.7, linetype="dashed" ) + geom_hline(aes(yintercept=2.1),col="yellow
")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



In conclusion, based on what we have plot, we found the life expectancy and GDP keep increasing but the birth rate decrease at the same time. Therefore, we can say that people's life condition is better then several decades ago and they can live longer. However, more and more people are unwilling to give birth in recent years.