ma615_midterm project

xiaofan xia

2019/10/15

Introduction

The dataset is from the world bank open data. This is a dataset of the health nutrition and population. It contains 259 countries and 59 years with 10 series of variables including adolescent fertility rate (births per 1,000 women ages 15-19), adults (ages 15+) and chidren (ages 0-14) living with HIV, adults and children newly infected with HIV, adults living with HIV, adults newly infected with HIV, age at first marriage (female), age at first marriage (male), age dependency ratio (% of working-age population), age dependency ratio (old),and age dependency ratio (young).

Firstly, I read in the data and clean the data. I specifically choose the variable of adolescent fertility rate (birth per 1,000 women ages 15-19). The cleaned data contains 4 variables and more than 15,000 observations. I mainly focus on the relationship between the adolescent fertility rate and the country areas as well as the relationship between the adolescent fertility rate and the country income levels.

```
data <- read.csv("data2.csv",encoding = "UTF-8")
year <- colnames(data)[5:63]
data_new <-
  data %>%
  gather(year,key = "year",value = "value")%>%
  select(Series.Name,Country.Name,year,value)
data_new$year <- as.numeric(str_sub(data_new$year,2,5))
data_new$value <- as.numeric(data_new$value)

data_adolscent_fert_rate <-
  data_new%>%
  filter(Series.Name=="Adolescent fertility rate (births per 1,000 women ages 15-19)")
head(tibble(data_adolscent_fert_rate))
```

```
## # A tibble: 6 x 1
##   data_adolscent_fert_rate$Series.Name           $Country.Name $year $value
##   <fct>                                          <fct>         <dbl>  <dbl>
## 1 Adolescent fertility rate (births per 1,000 w~ Afghanistan    1960  145.
## 2 Adolescent fertility rate (births per 1,000 w~ Albania        1960   54.4
## 3 Adolescent fertility rate (births per 1,000 w~ Algeria        1960  124.
## 4 Adolescent fertility rate (births per 1,000 w~ American Sam~  1960   NA
## 5 Adolescent fertility rate (births per 1,000 w~ Andorra        1960   NA
## 6 Adolescent fertility rate (births per 1,000 w~ Angola         1960  202.
```

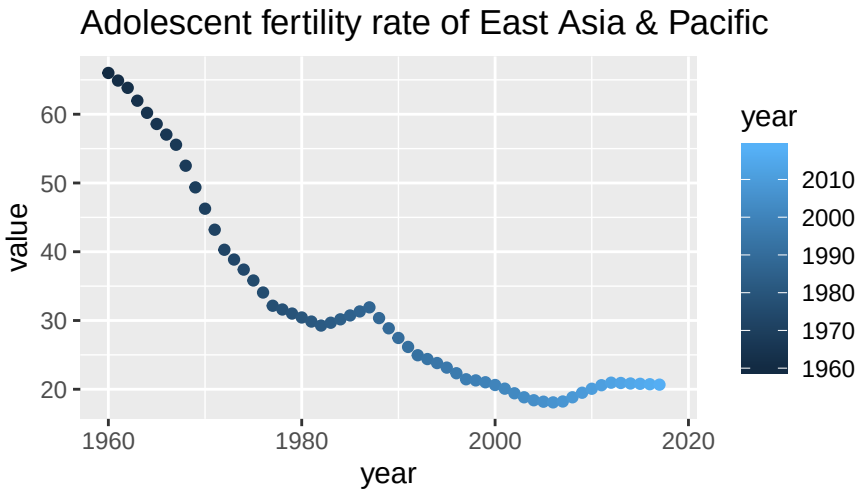
EDA by world areas.

First, I want to dig into the data about the adolescent fertility rate by explore the different relationship between the rate and the time in different world areas such as European and Central Asia and Arab World.

Therefore, I plot the adolescent fertility rate by year for each world area, and then I compare them in one same plot. There are some interesting facts I find from the plot. For instance, there are even some increases at some particular time point and I believe it must relate to some world wide events at that time point somehow.

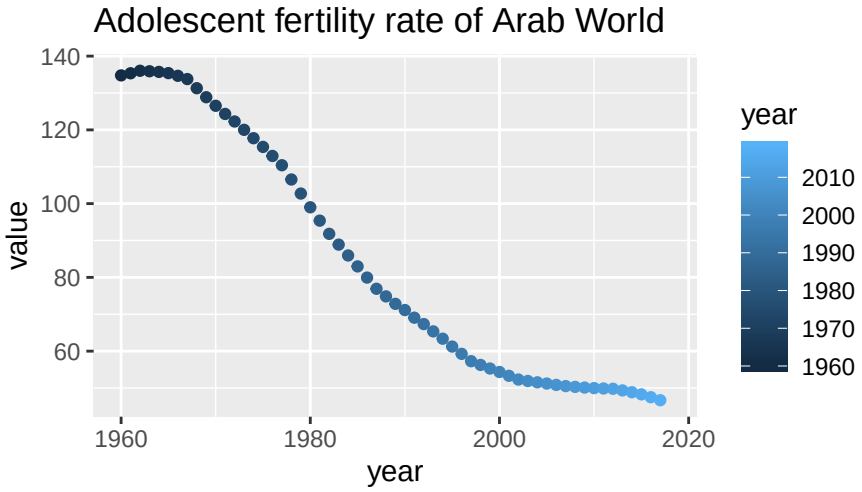
East Asia and Pacific Area

```
East_Asia_Pacific <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="East Asia & Pacific")
ggplot(East_Asia_Pacific)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate
```



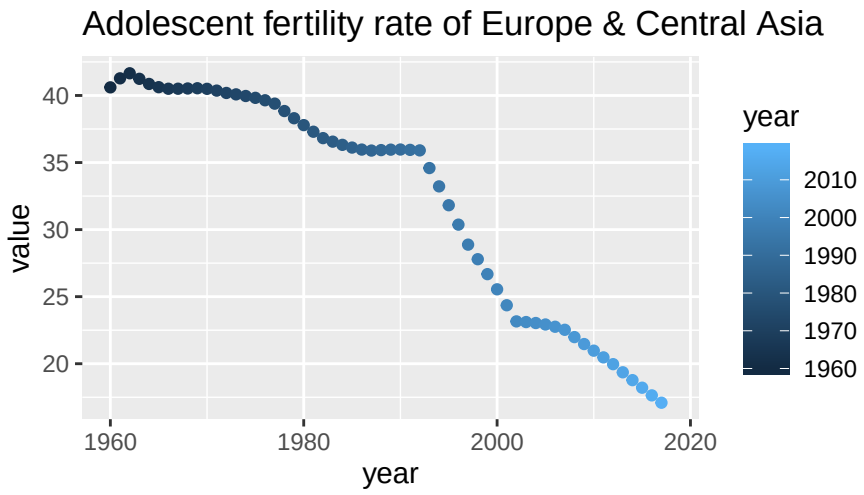
Arab World Area

```
Arab_World <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Arab World")
ggplot(Arab_World)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of Ara
```



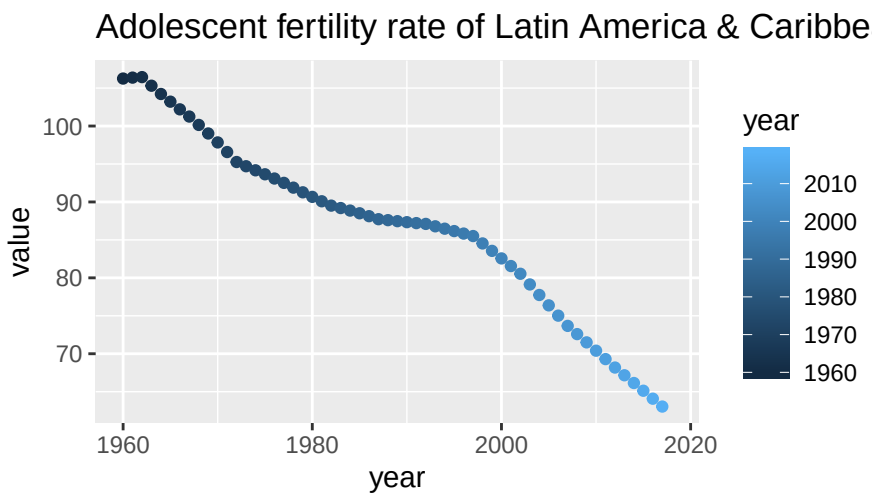
Europe and Central Asia Area

```
Europe_Central_Asia <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Europe & Central Asia")
ggplot(Europe_Central_Asia)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility ra
```



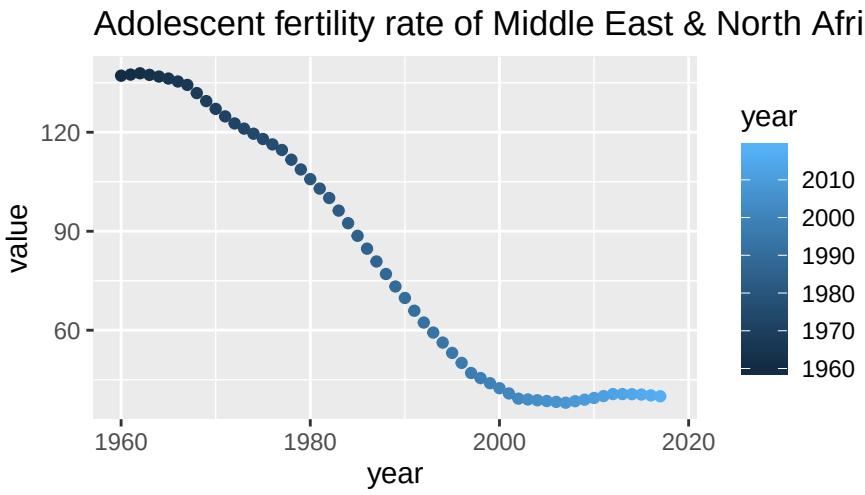
Latin America and Caribbean Area

```
Latin_America_Caribbean <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Latin America & Caribbean")
ggplot(Latin_America_Caribbean)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertili
```



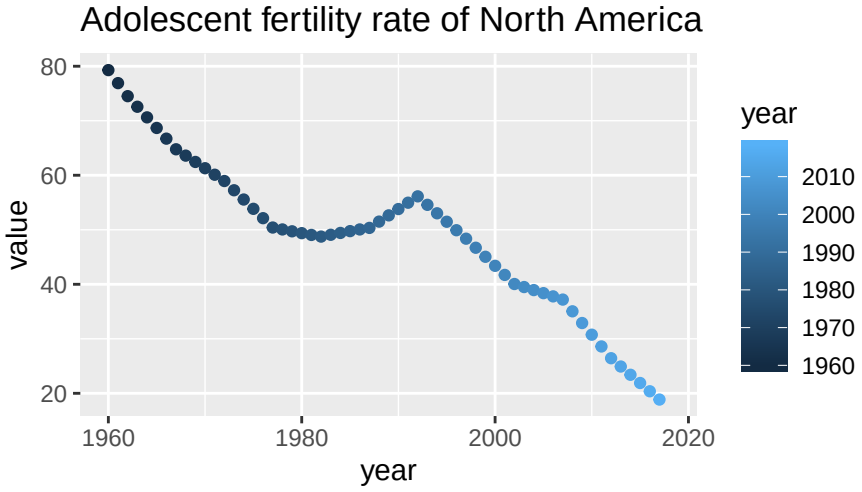
Middle East and North Africa Area

```
Middle_East_North_Africa <-
 data_adolscent_fert_rate%>%
  filter(Country.Name=="Middle East & North Africa")
ggplot(Middle_East_North_Africa)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertili
```



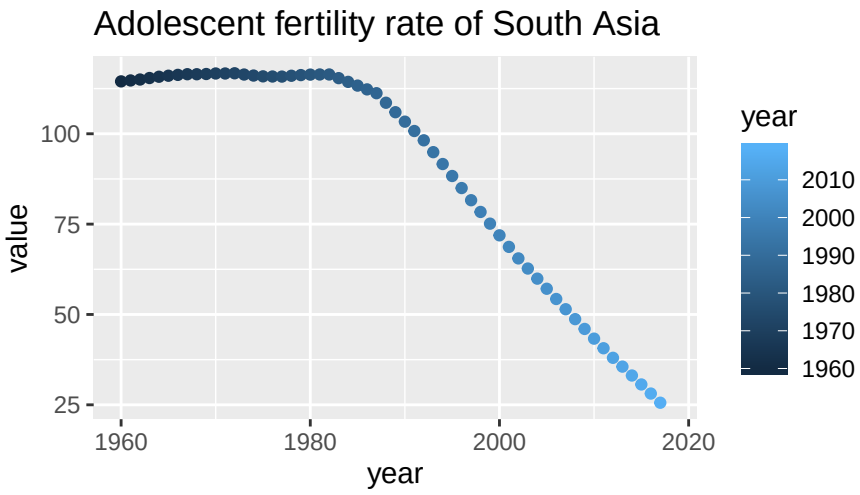
North America Area

```
North_America <-
   data_adolscent_fert_rate%>%
  filter(Country.Name=="North America")
ggplot(North_America)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of
```



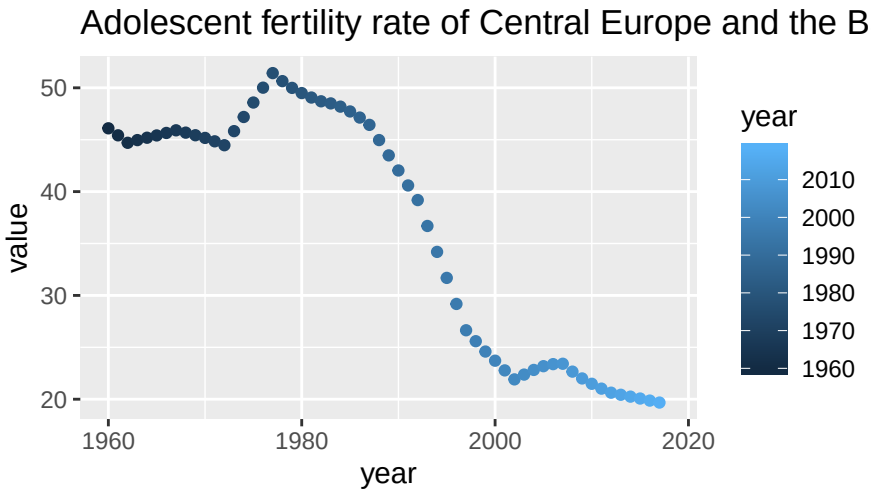
South Asia Area

```
South_Asia <-
   data_adolscent_fert_rate%>%
  filter(Country.Name=="South Asia")
ggplot(South_Asia)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of Sou
```



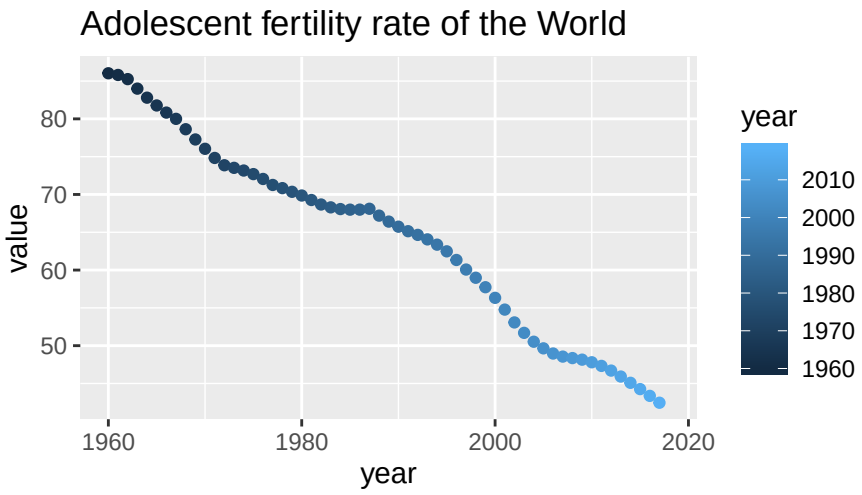
Central Europe and the Baltics Area

```
Central_Europe_the_Baltics <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Central Europe and the Baltics")
ggplot(Central_Europe_the_Baltics)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent ferti
```



The Whole World

```
World <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="World")
ggplot(World)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of the Worl
```

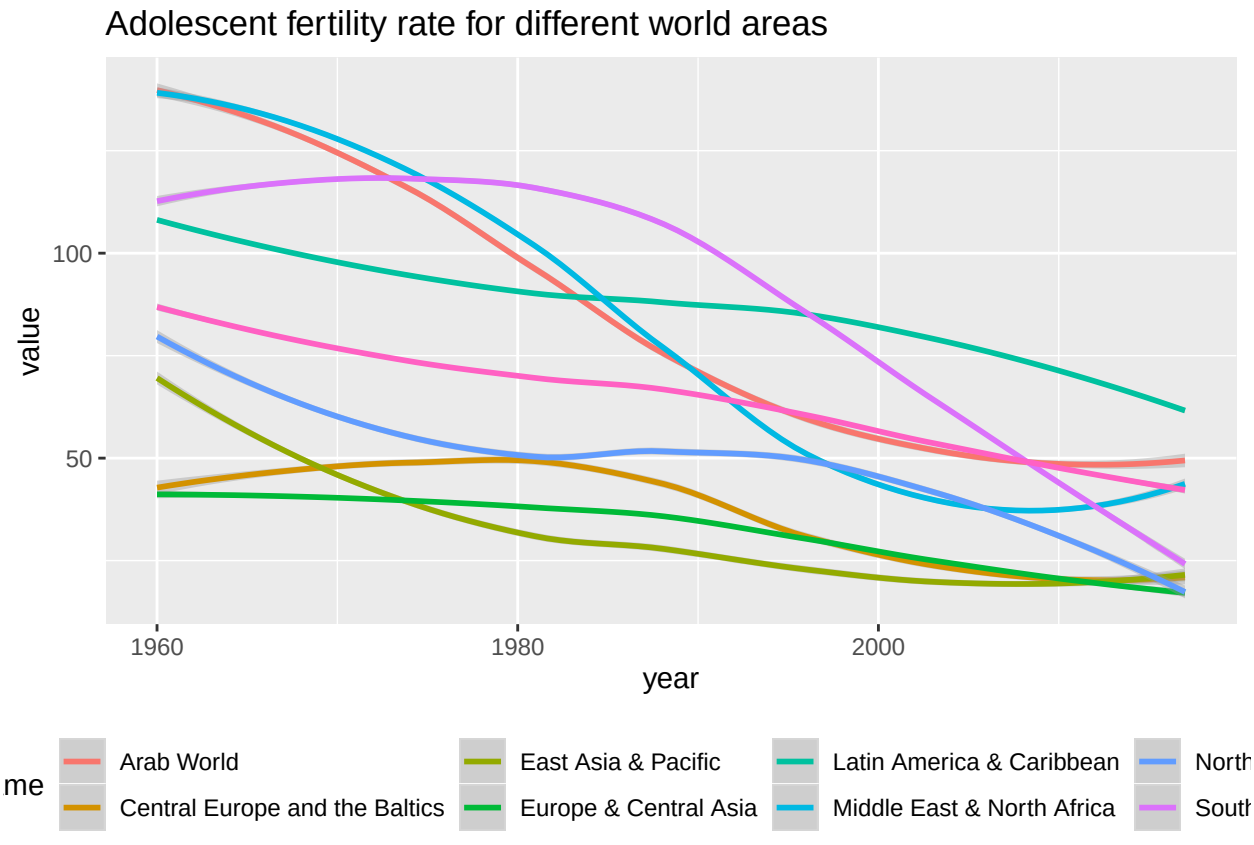


Different Areas in Total

```
region <- rbind(World,Central_Europe_the_Baltics,South_Asia,North_America,Middle_East_North_Africa,Latin

ggplot(region)+aes(year,value)+geom_smooth(aes(color=Country.Name))+ggtitle("Adolescent fertility rate f
```

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




```
#show differences regarding to different world areas
```

In the last plot, we can see that the total rate is on downward trend, but there are still some differences in the trend patterns. Some areas have faster decreasing rate than others. Overall, the adolescent fertility rate

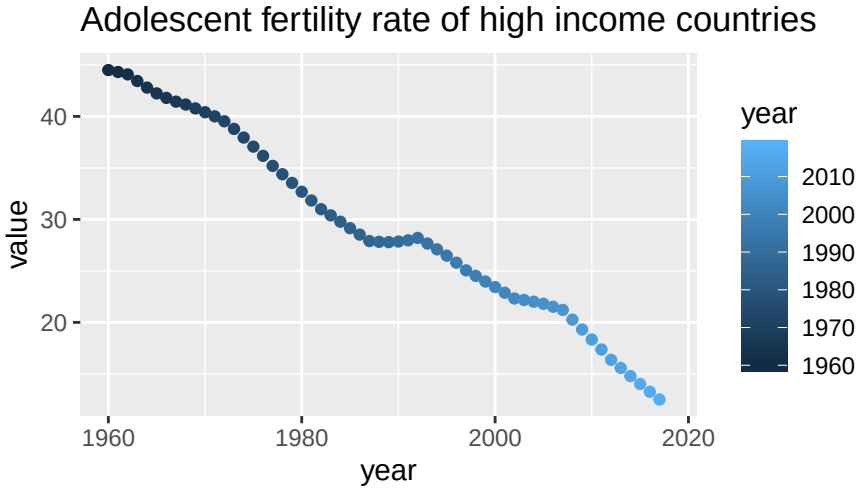
decreases gradually from 1960 to 2018. But the patterns of different areas in the world is quite different. This might also reflect the relationship between economic development and adolescent fertility rate as well. Therefore, I want to explore the adolescent fertility rate differences in different income levels.

EDA by countries' income level.

I plot the adolescent fertility rate by year for each group of countries with differnt income levels, and then I compare them in one same plot.

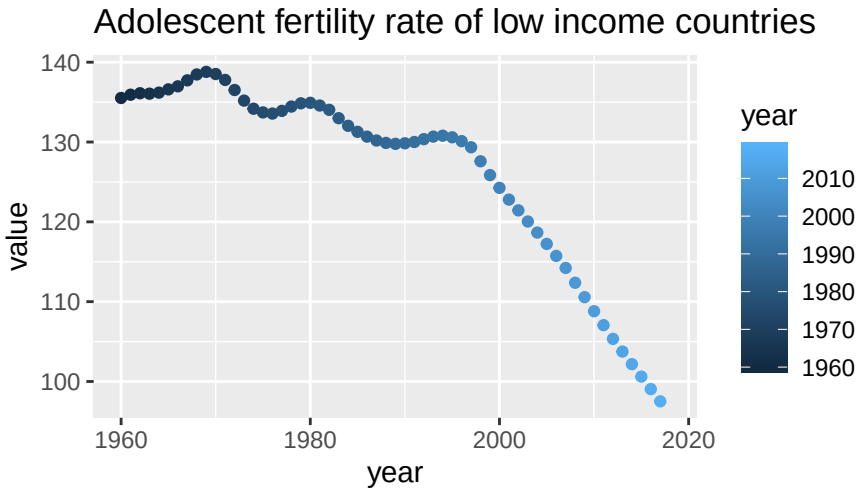
High Income Countries

```
high_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="High income")
ggplot(high_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of hi
```



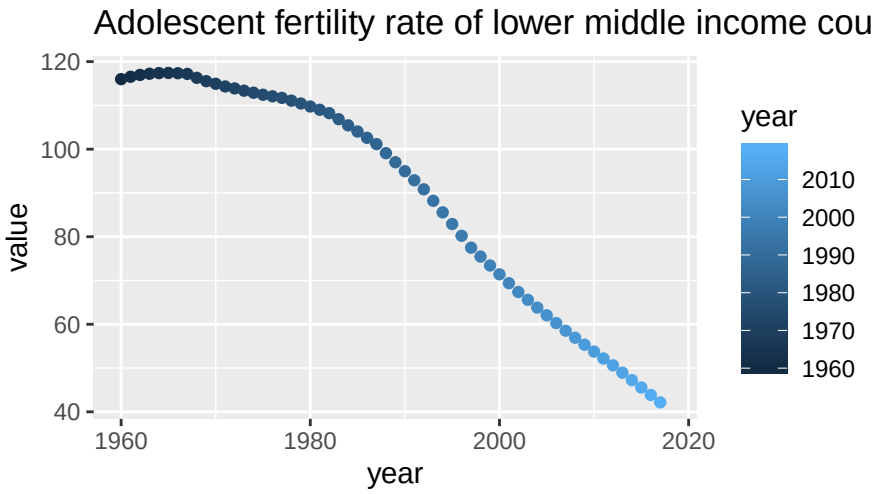
Low Income Countries

```
low_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Low income")
ggplot(low_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of lo
```



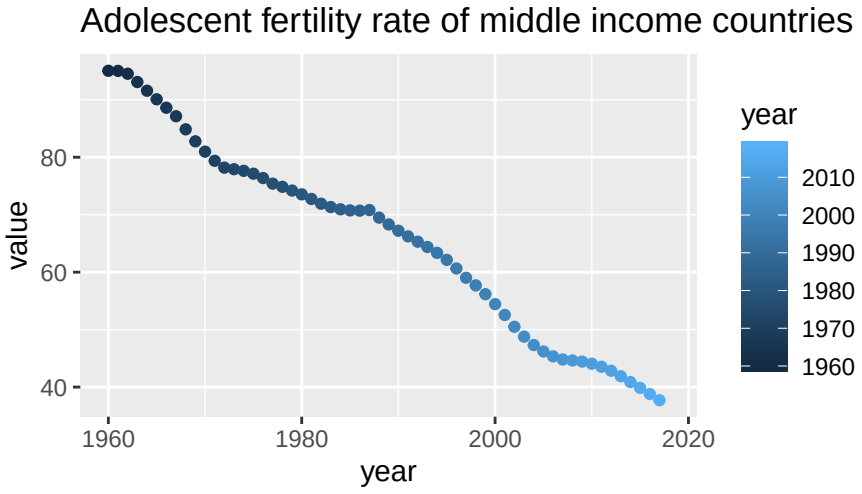
Lower and Middle Income Countries

```
lower_middle_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Lower middle income")
ggplot(lower_middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility ra
```



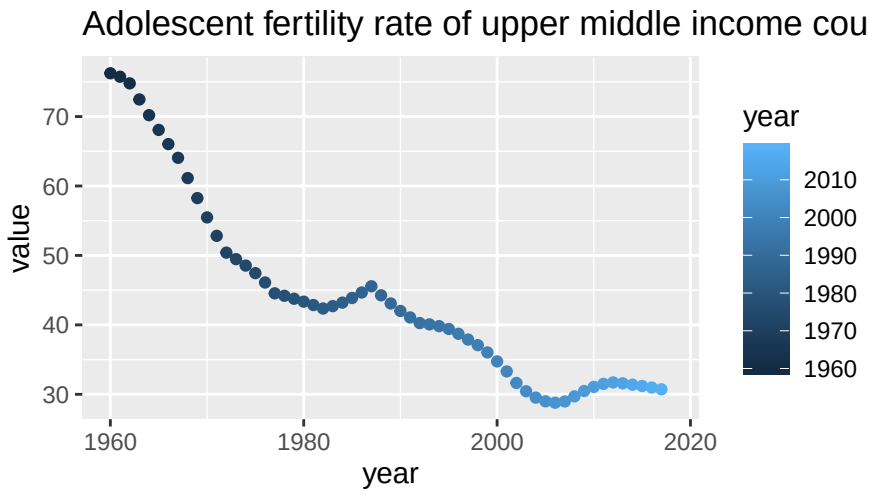
Middle Income Countries

```
middle_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Middle income")
ggplot(middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of
```



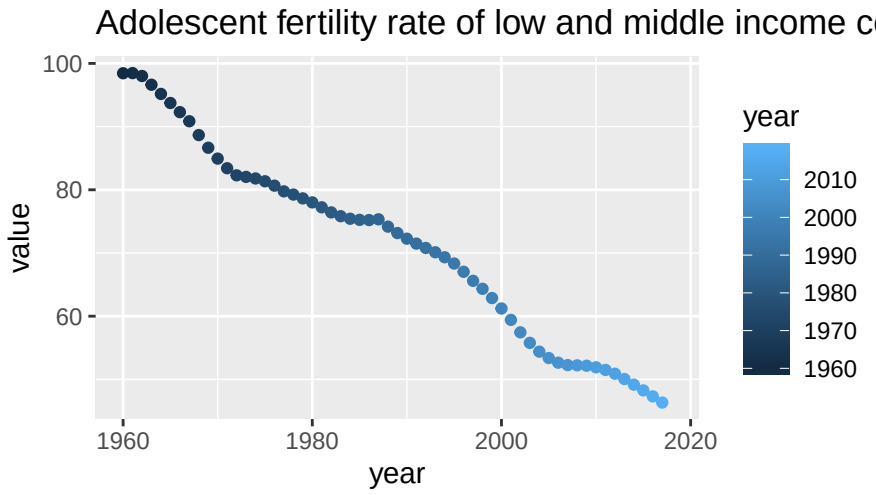
Upper and Middle Income Countries

```
upper_middle_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Upper middle income")
ggplot(upper_middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility ra
```



Low and Middle Income Countries

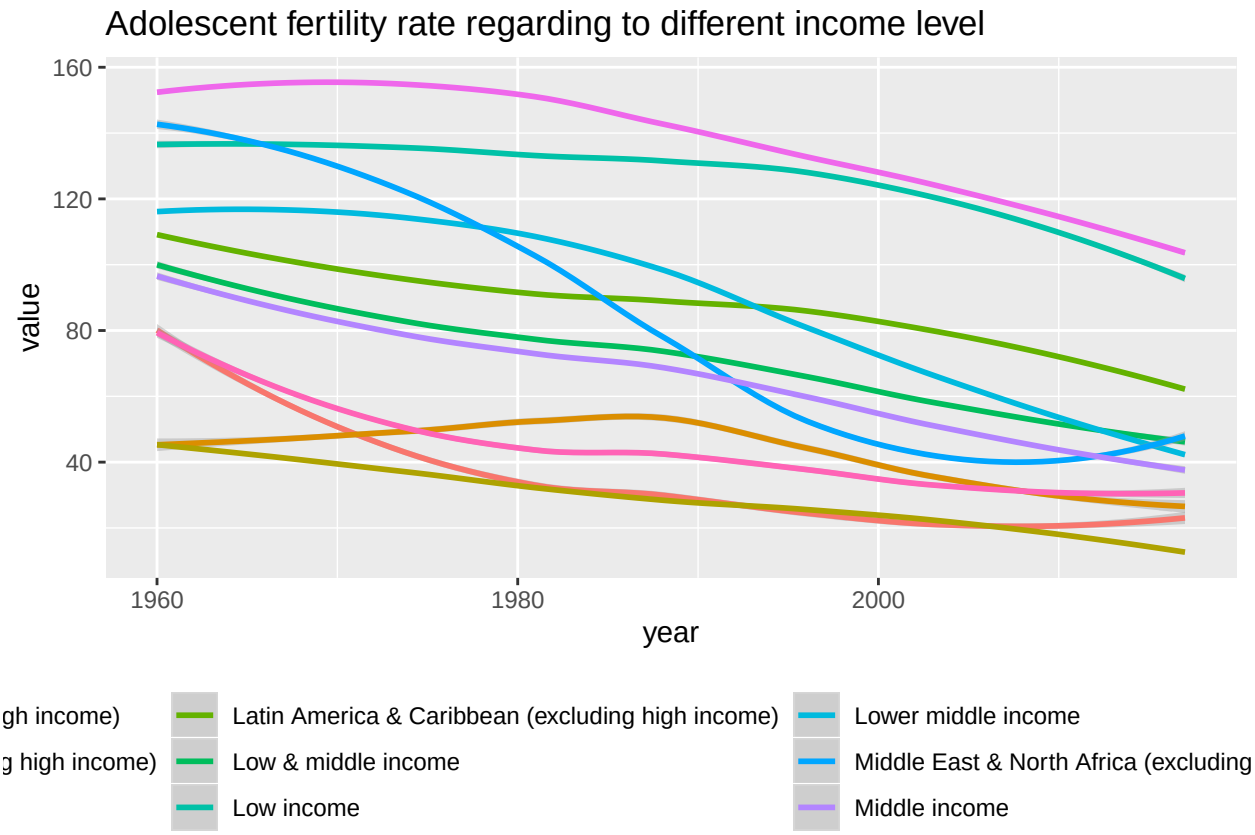
```
low_middle_income <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Low & middle income")
ggplot(low_middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rat
```



Countries with different income levels in total

```
income <-
  data_adolscent_fert_rate%>%
  filter(grepl("income",data_adolscent_fert_rate$Country.Name)==TRUE)
ggplot(income)+aes(year,value)+geom_smooth(aes(color=Country.Name))+ggtitle("Adolescent fertility rate
```

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




```
#compare the rate regarding to different income levels
```

It is pretty obvious that low income countries started the decreasing of adolescent fertility rate later than countries with higher income level.

Conclusion

Overall, the adolescent fertility rate is on a downward trend globally, falling year by year. Regarding to different areas of the world and different income levels, the trends vary a little bit. The rates of higher-income countries fall down sharply from 1960 to 1980, but lower-income countries generally fall down a lot after 1980. As for different regions, European and North America is quite ahead on decreasing the rates. But generally, the whole world is gradually dimishing the adolescent ferility rate. Decreased adolescent fertility rate means improvement in health care services as well as in other related fields such as education levels. Hopefully we can keep decrease the adolescent fertility rate gradually in the future.