

# ma615\_midterm project

*xiaofan xia*

*2019/10/15*

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 children (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).

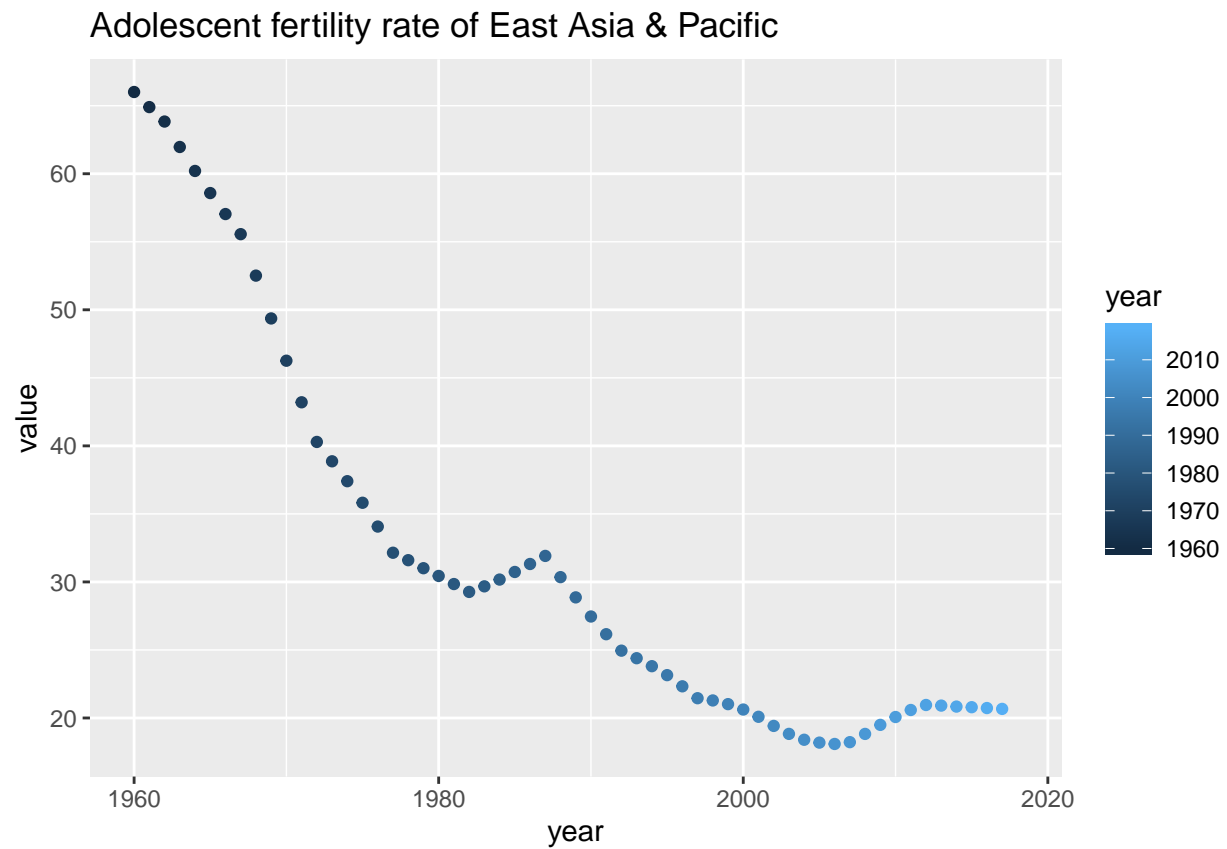
Firstly, I read in the data and clean the data. I specifically choose the variable of adolescent fertility rate (birth per 1,000 women ages 15-19).

```
data <- read.csv("data2.csv")
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)

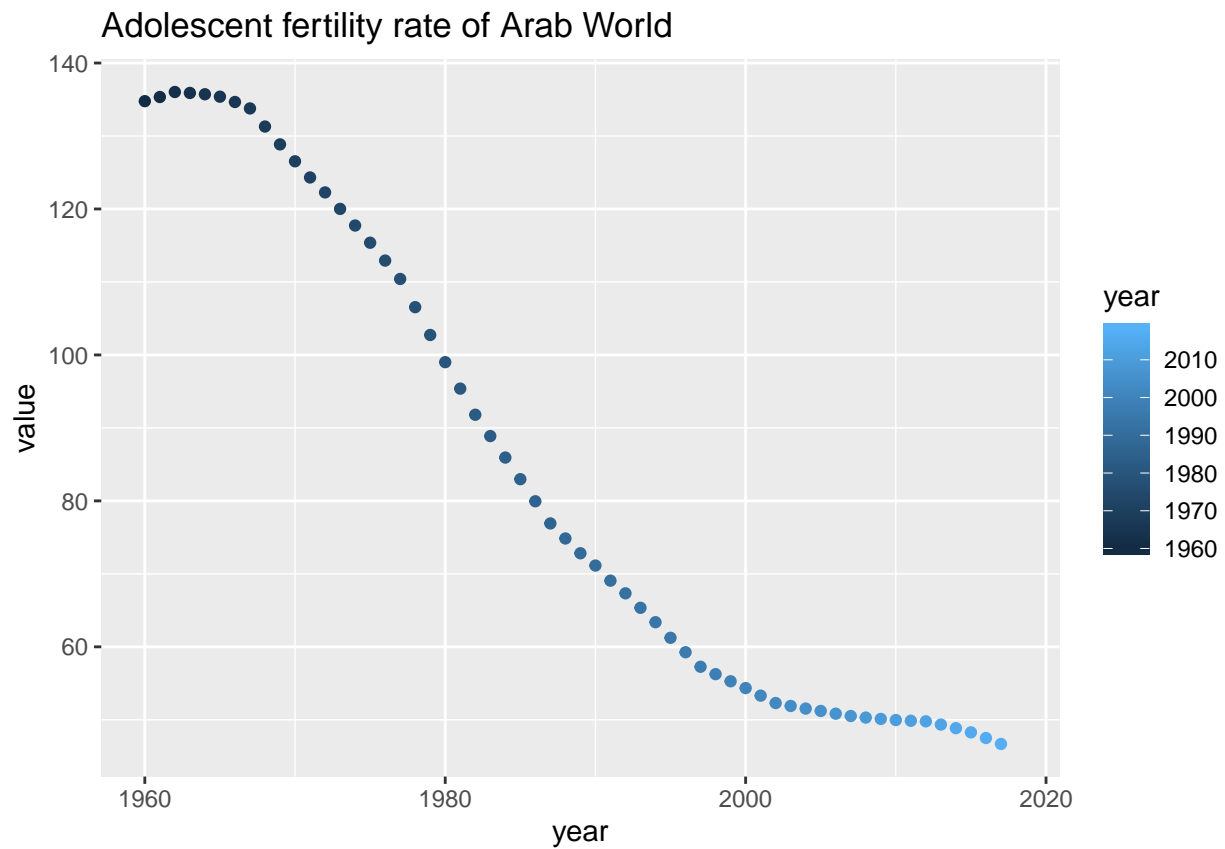
data_adolescent_fert_rate <-
  data_new %>%
  filter(Series.Name == "Adolescent fertility rate (births per 1,000 women ages 15-19)")
```

I did EDA by world areas.

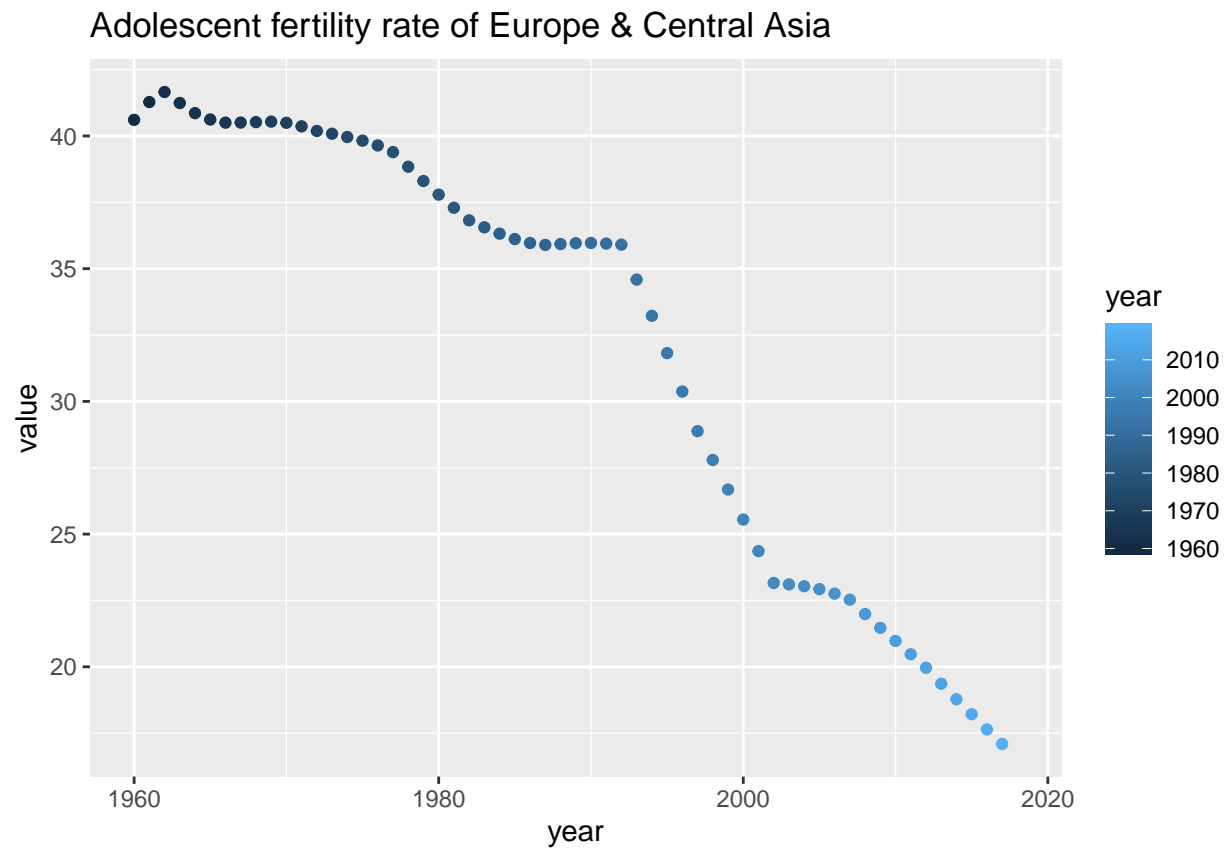
```
East_Asia_Pacific <-
  data_adolescent_fert_rate %>%
  filter(Country.Name == "East Asia & Pacific")
ggplot(East_Asia_Pacific) + aes(year, value) + geom_point(aes(color=year)) + ggtitle("Adolescent fertility rate")
```



```
Arab_World <-
  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 Arab World")
```

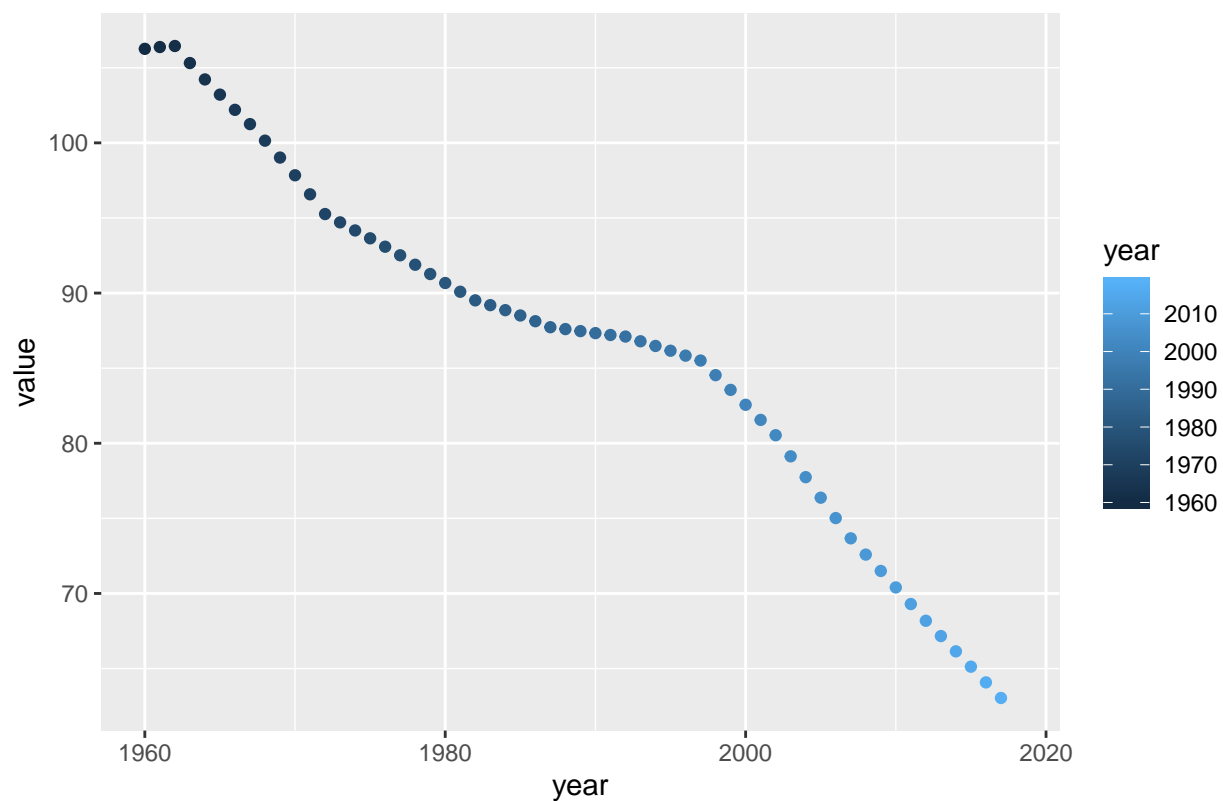


```
Europe_Central_Asia <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Europe & Central Asia")
ggplot(Europe_Central_Asia)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility r
```



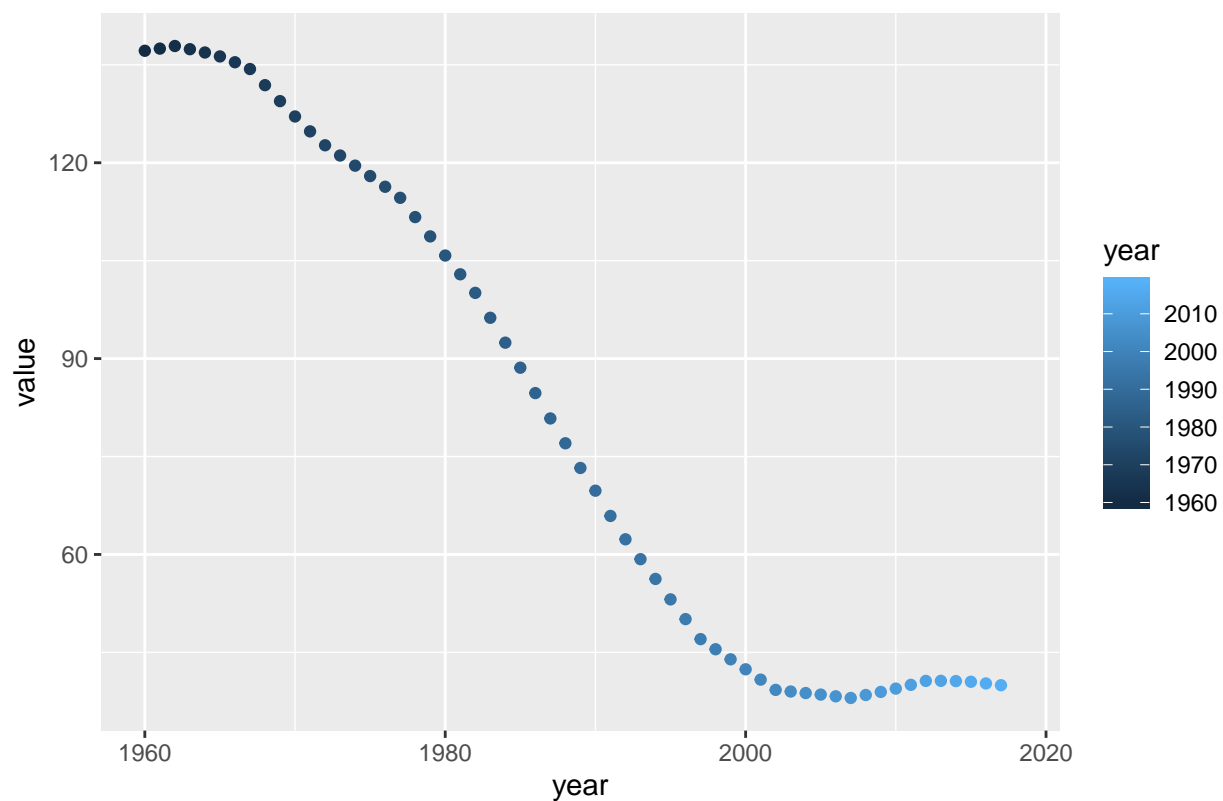
```
Latin_America_Caribbean <-  
  data_adolscent_fert_rate%>%  
  filter(Country.Name=="Latin America & Caribbean")  
ggplot(Latin_America_Caribbean)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertili
```

## Adolescent fertility rate of Latin America & Caribbean

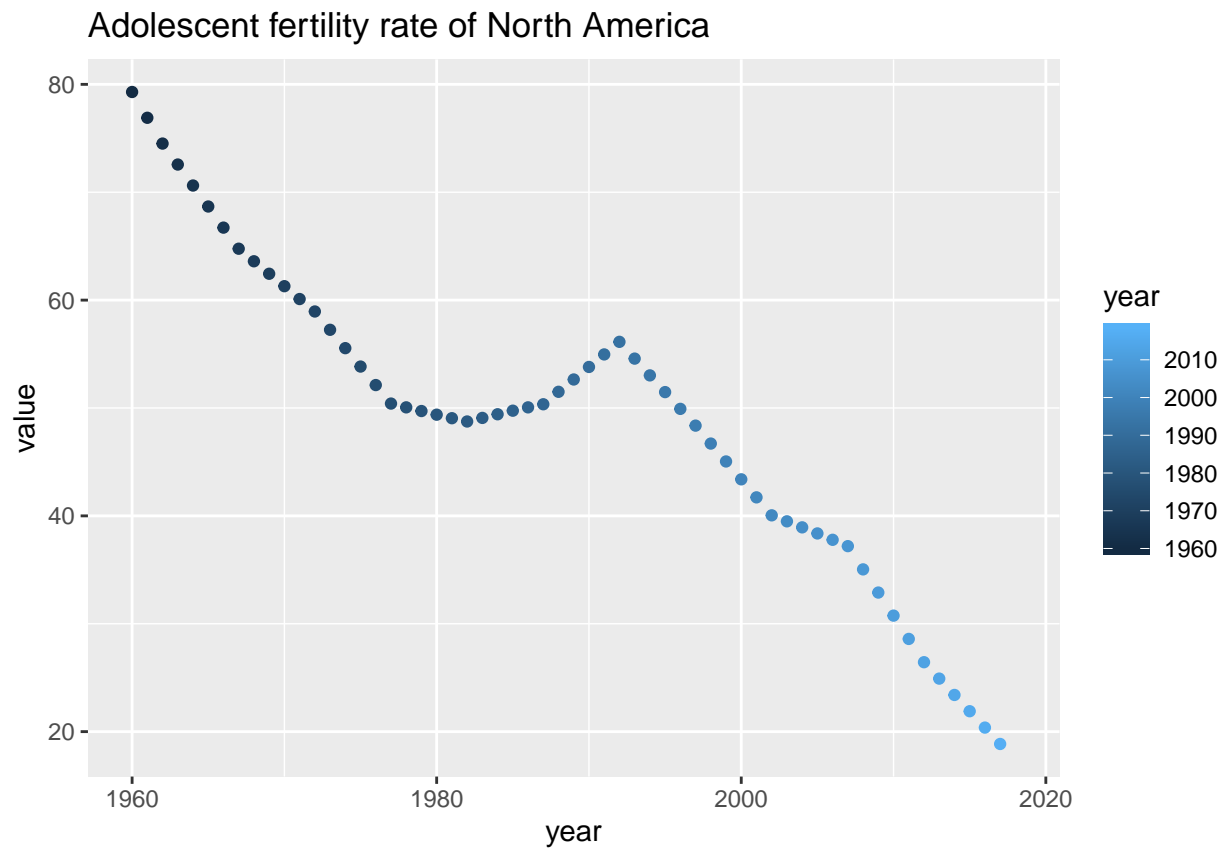


```
Middle_East_North_Africa <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Middle East & North Africa")
ggplot(Middle_East_North_Africa)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent ferti
```

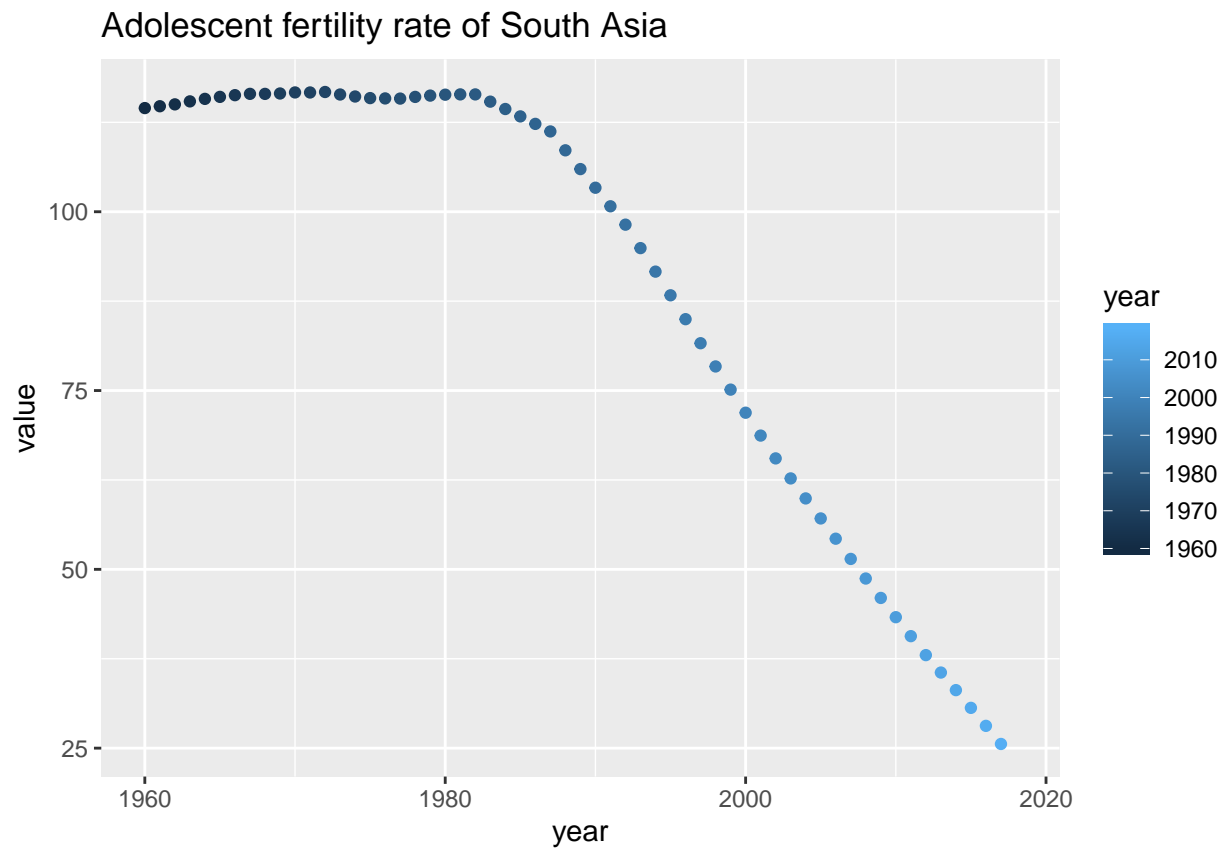
## Adolescent fertility rate of Middle East & North Africa



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

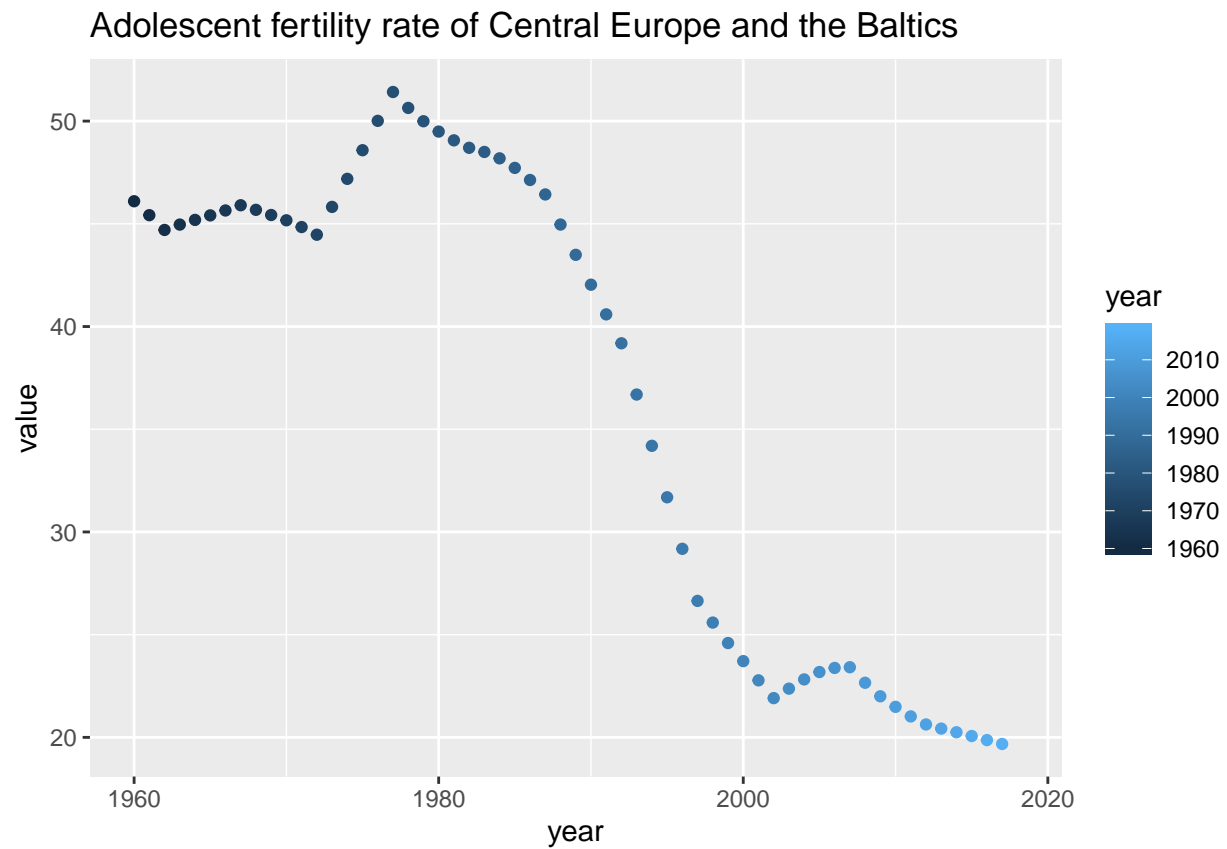


```
South_Asia <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="South Asia")
ggplot(South_Asia)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of South Asia")
```

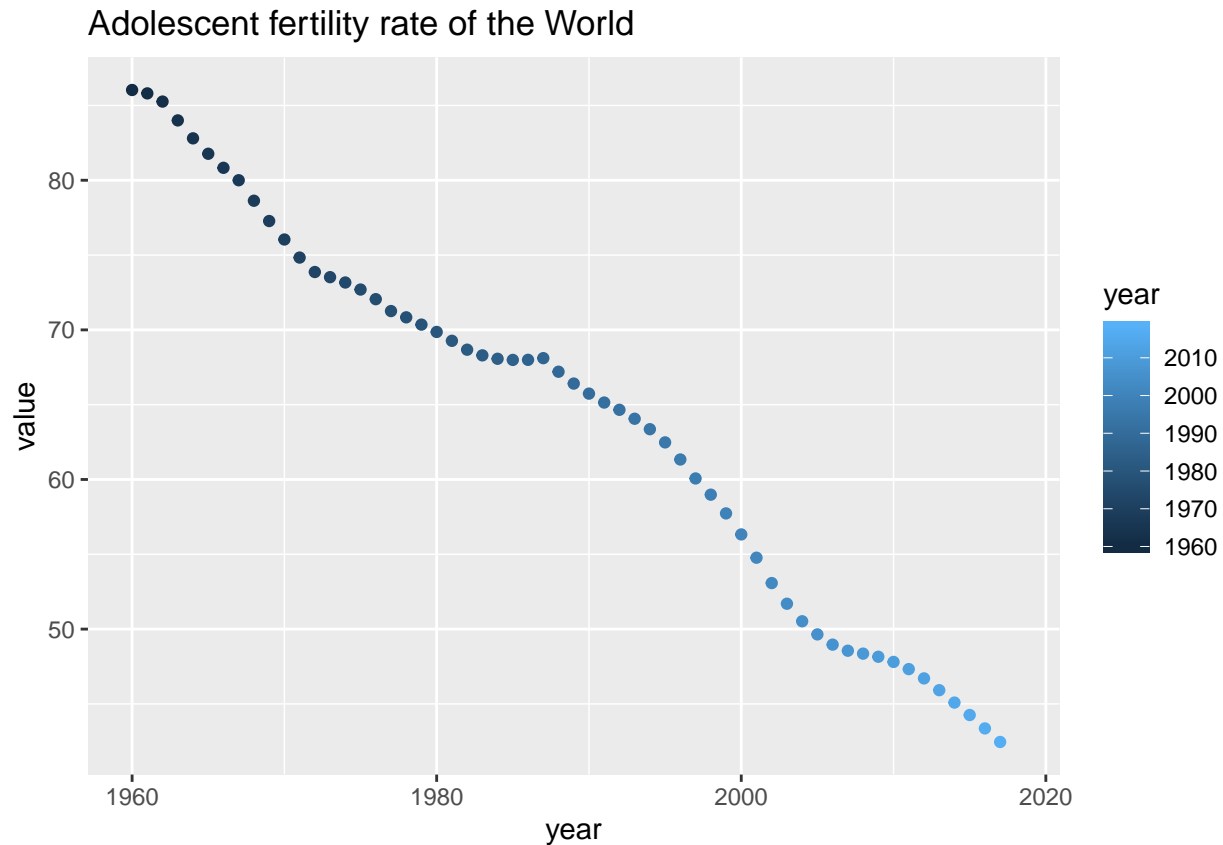


```
Central_Europe_the_Baltics <-
  data_adolscent_fert_rate%>%
  filter(Country.Name=="Central Europe and the Baltics")
ggplot(Central_Europe_the_Baltics)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fert
```





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

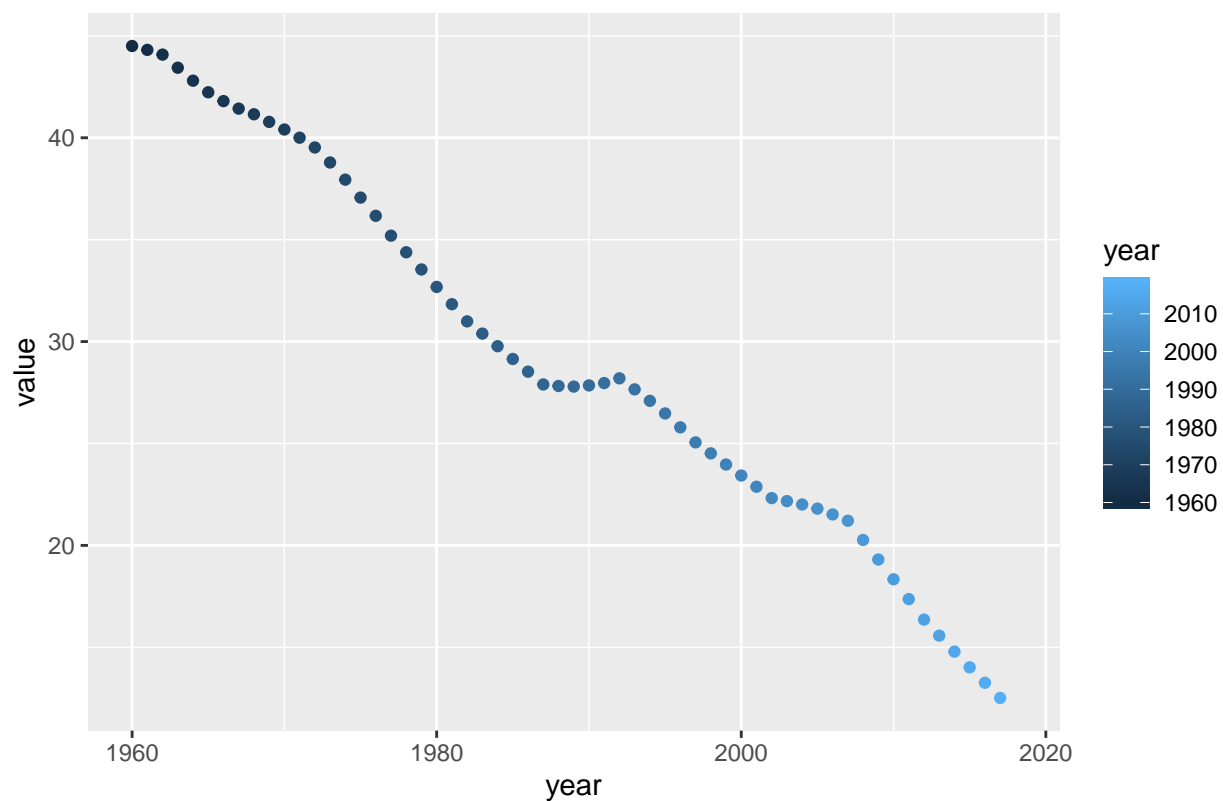


Overall, the adolescent fertility rate decreases gradually from 1960 to 2018. But the patterns of different areas in the world is quite different. This might also reflect the relationship between economic development and adolescent fertility rate as well. Therefore, I want to explore the adolescent fertility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

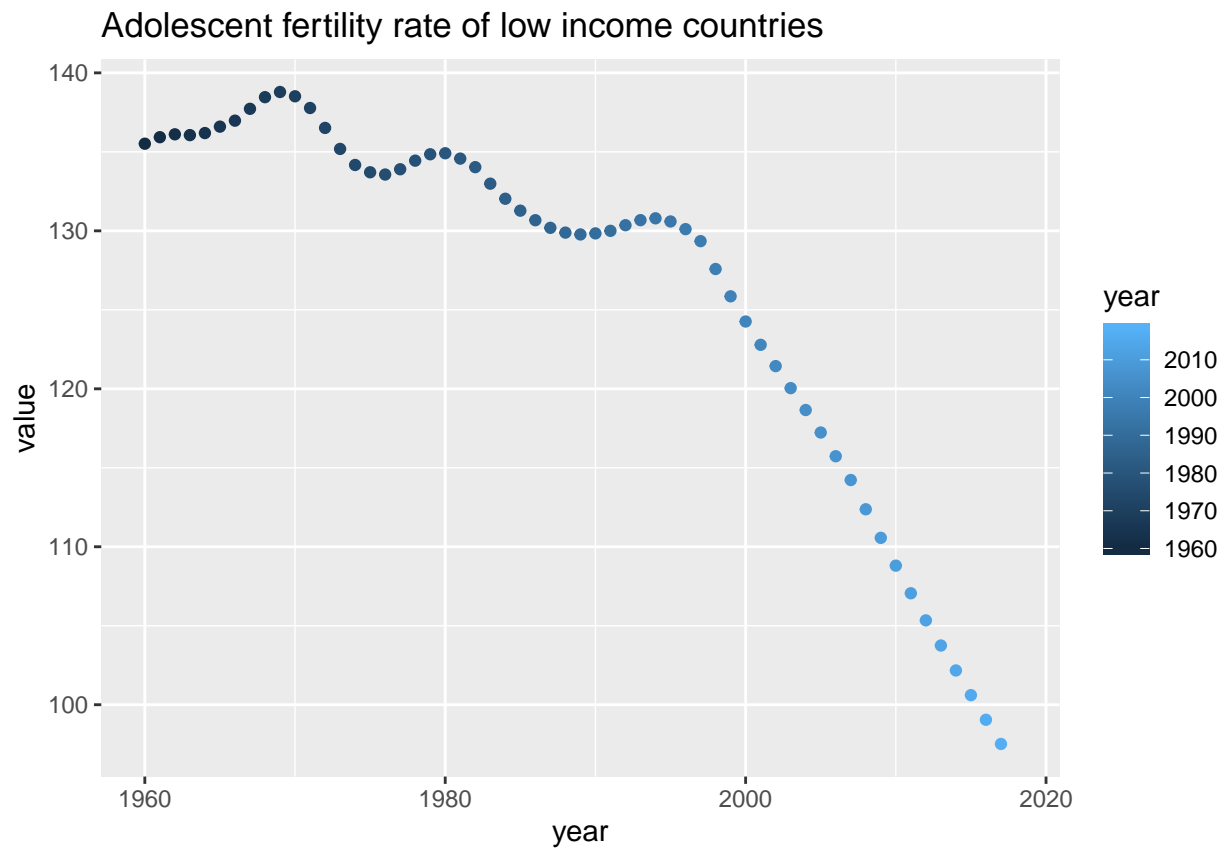
EDA by countries' income level.

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

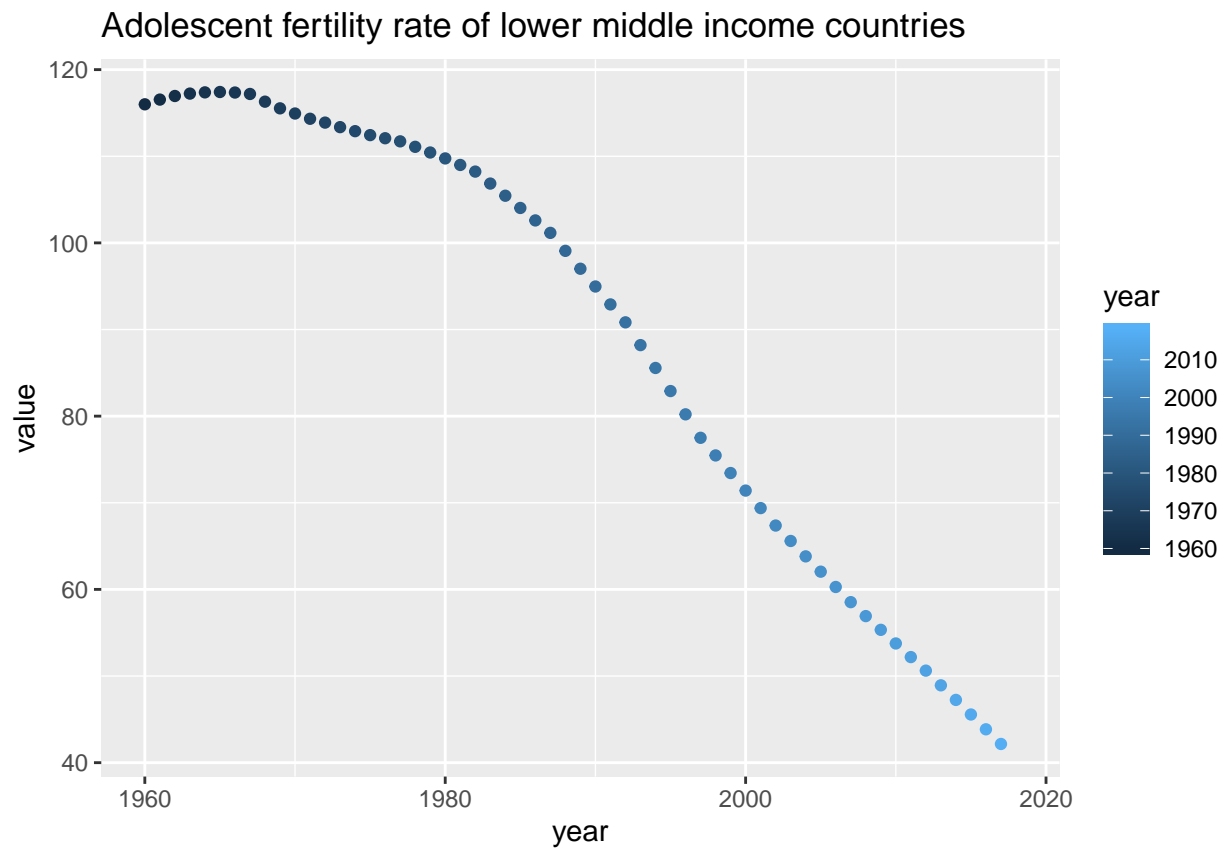
Adolescent fertility rate of high income countries



```
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 low income countries")
```

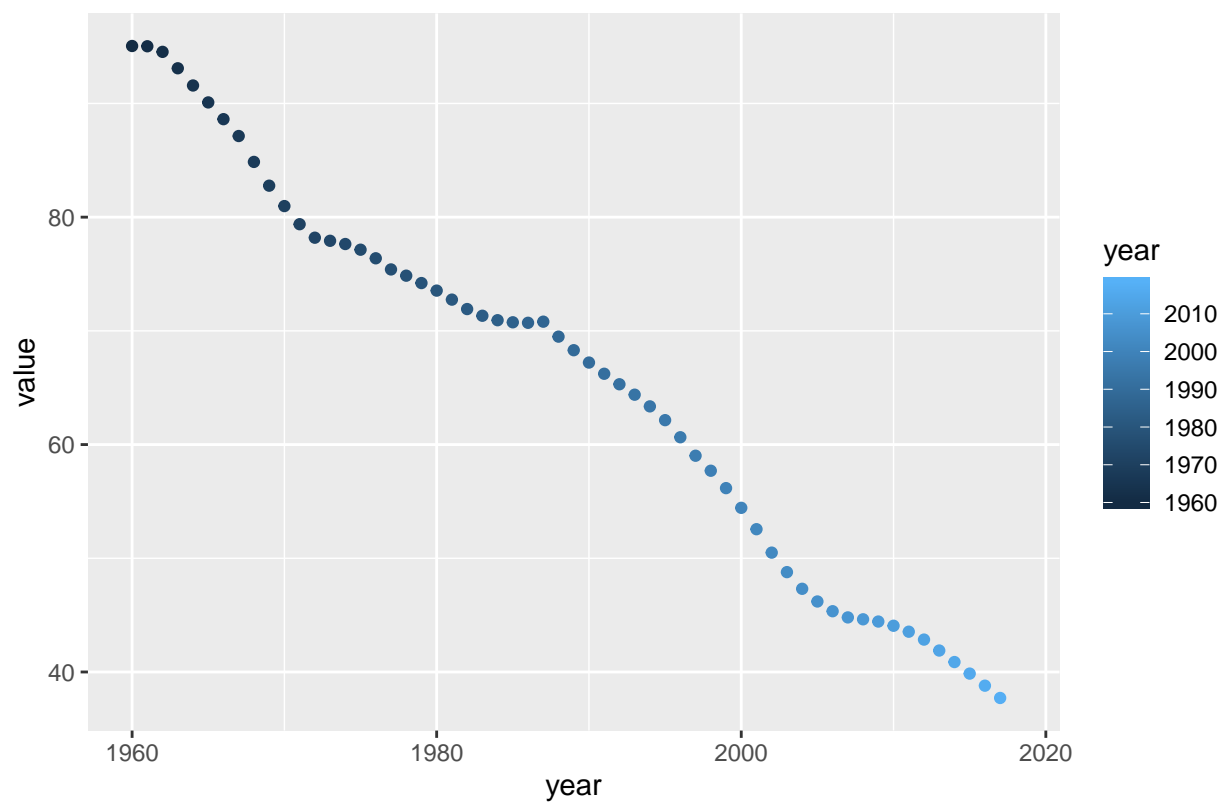


```
lower_middle_income <-  
  data_adolscent_fert_rate%>%  
  filter(Country.Name=="Lower middle income")  
ggplot(lower_middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility r
```



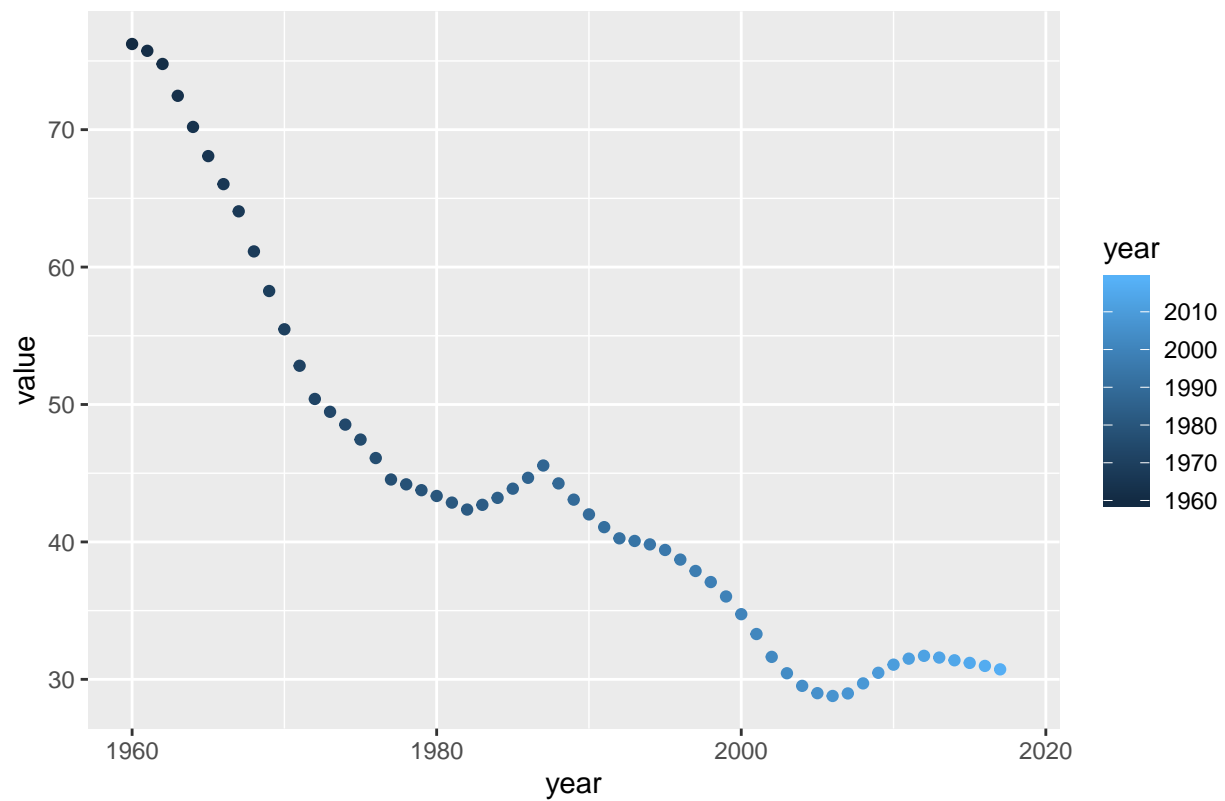
```
middle_income <-  
  data_adolscent_fert_rate%>%  
  filter(Country.Name=="Middle income")  
ggplot(middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility rate of
```

## Adolescent fertility rate of middle income countries

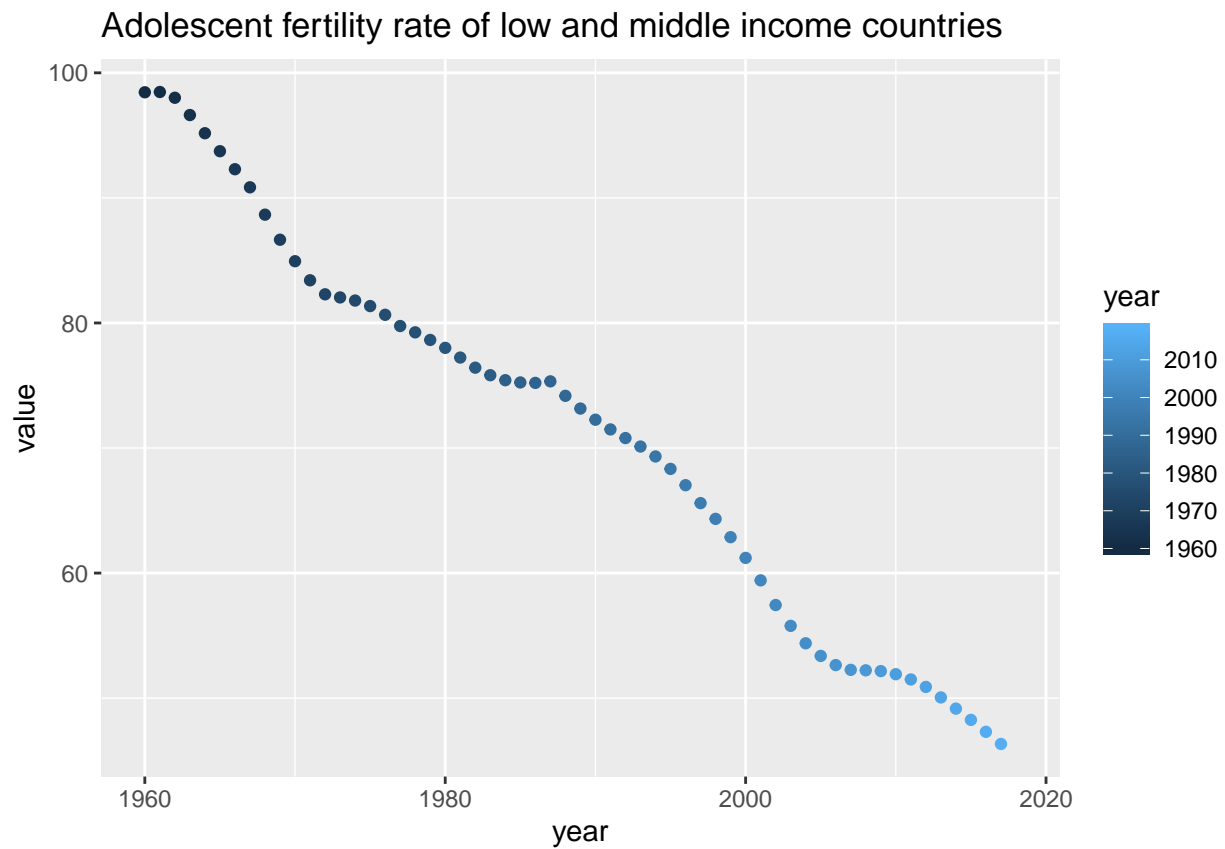


```
upper_middle_income <-  
  data_adolscent_fert_rate%>%  
  filter(Country.Name=="Upper middle income")  
ggplot(upper_middle_income)+aes(year,value)+geom_point(aes(color=year))+ggtitle("Adolescent fertility r
```

Adolescent fertility rate of upper 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 rate")
```



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