project homework

Chang KuiJian—-ZheJiang University

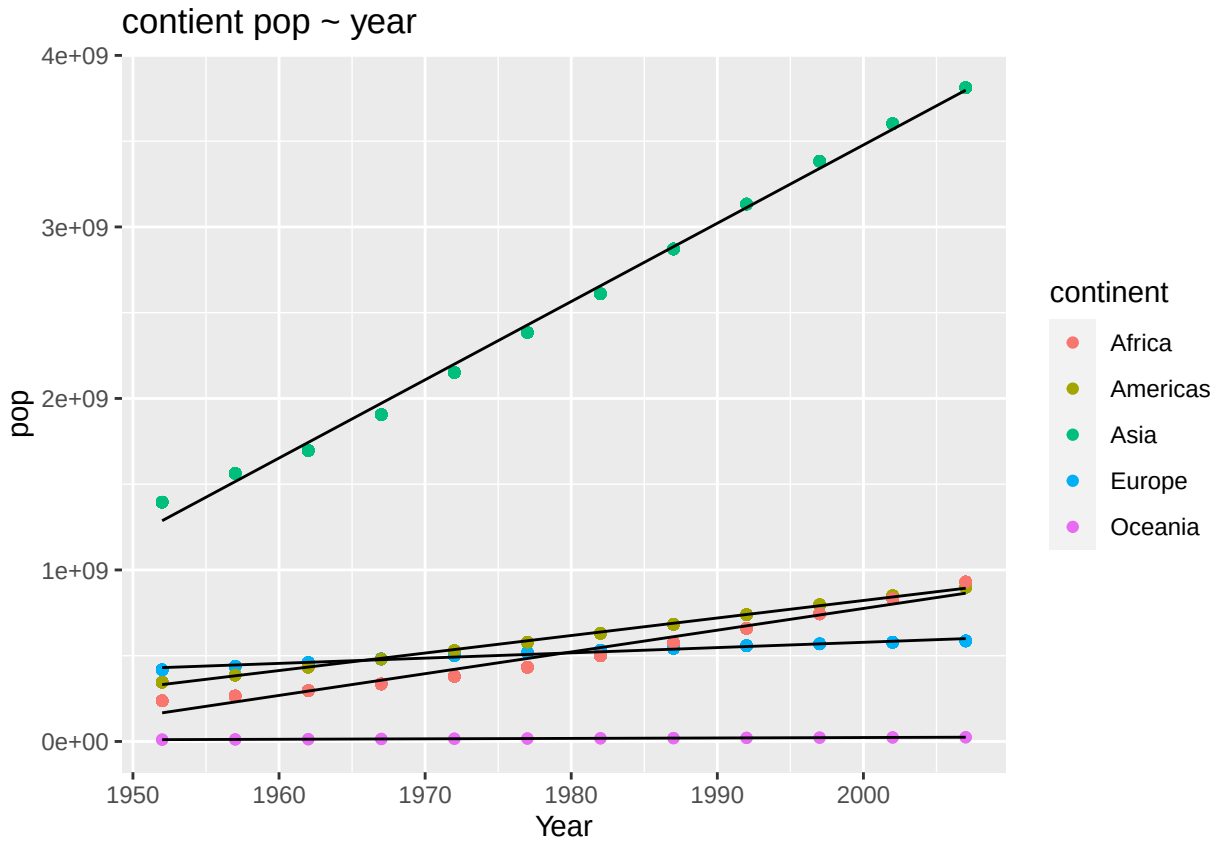
Abstract : A complete data analysis of the data *Gapminder* with R language using tidyverse, ggplot2, function, model etc.

Keyword : Gapminder , R Language , Data Analysis

Text :

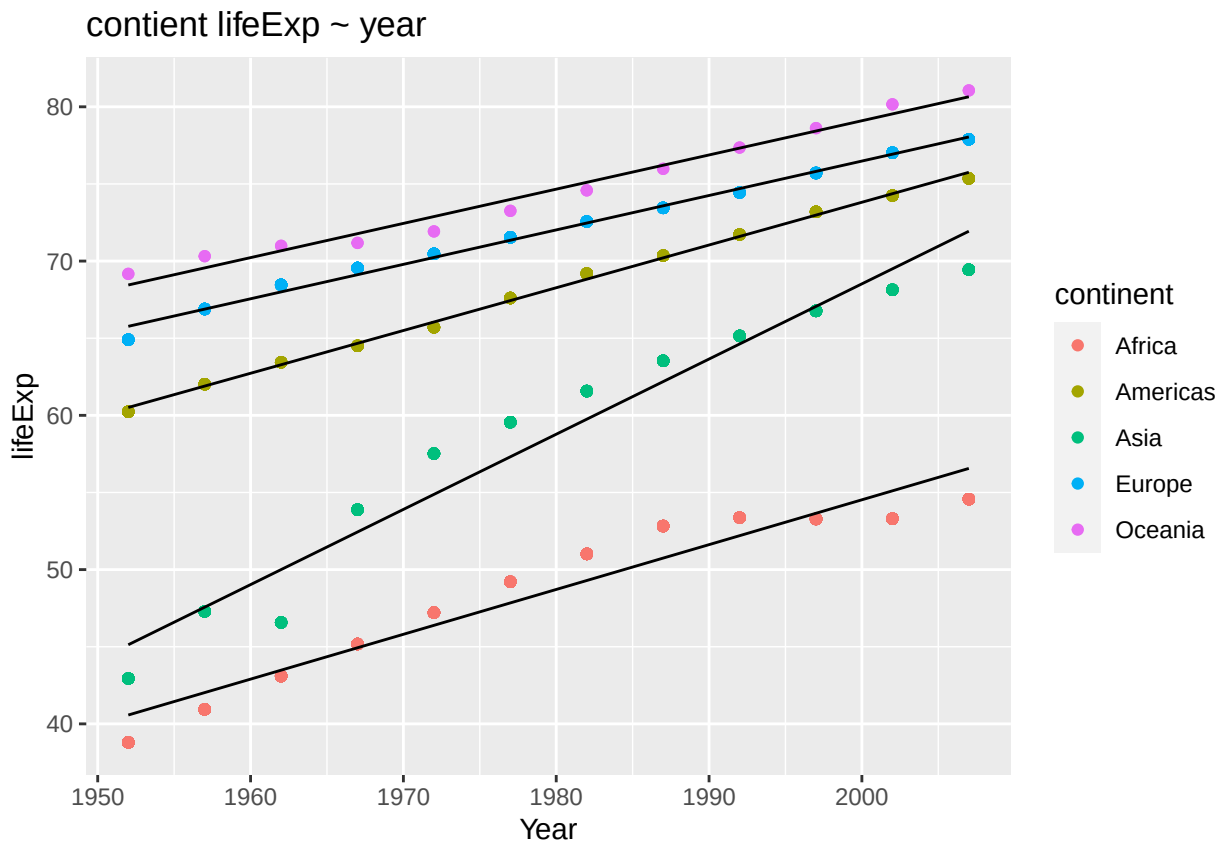
We will analyze the relationship between the variable that the gapminder.csv in data file contains or implies.We analyzed in four angles in total.

1.population of continent among years : We plot the population of each continent among years, and fit the point with linear model.According to the following plot, we conclude that the population of each continent has increased more or less since 1950s. All of them fit the linear model well, and the population of Asia has the fastest growing while that of Oceania has the slowest.

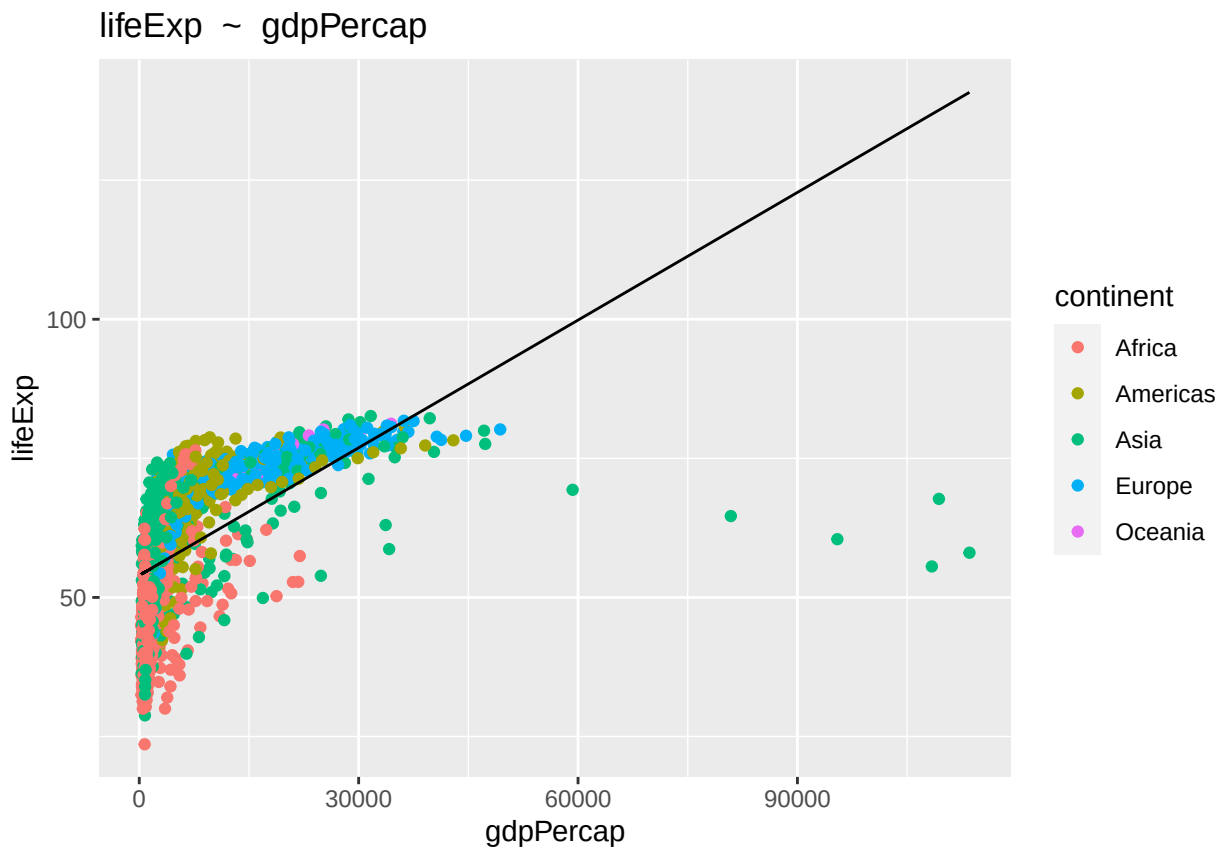


2.life expectation of continent among years : We plot the life expectation of each continent among years, and fit the point with linear model.According to the following plot, we conclude that the life expectation of

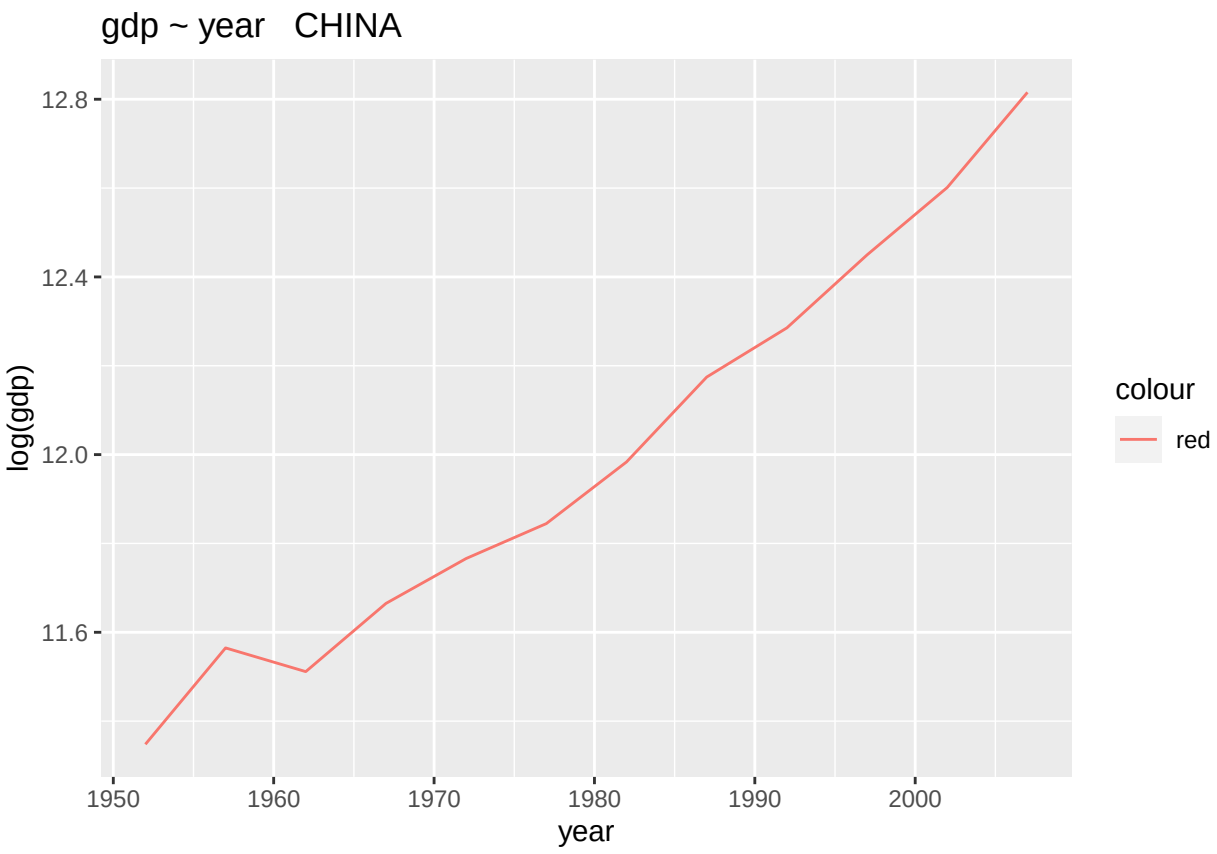
each continent has increased a lot since 1950s. All of them fit the linear model well, and the life expectation of Oceania has the highest value since 1950s while that of Africa has the lowest.But the growing rate of Asia and Africa are quite high.

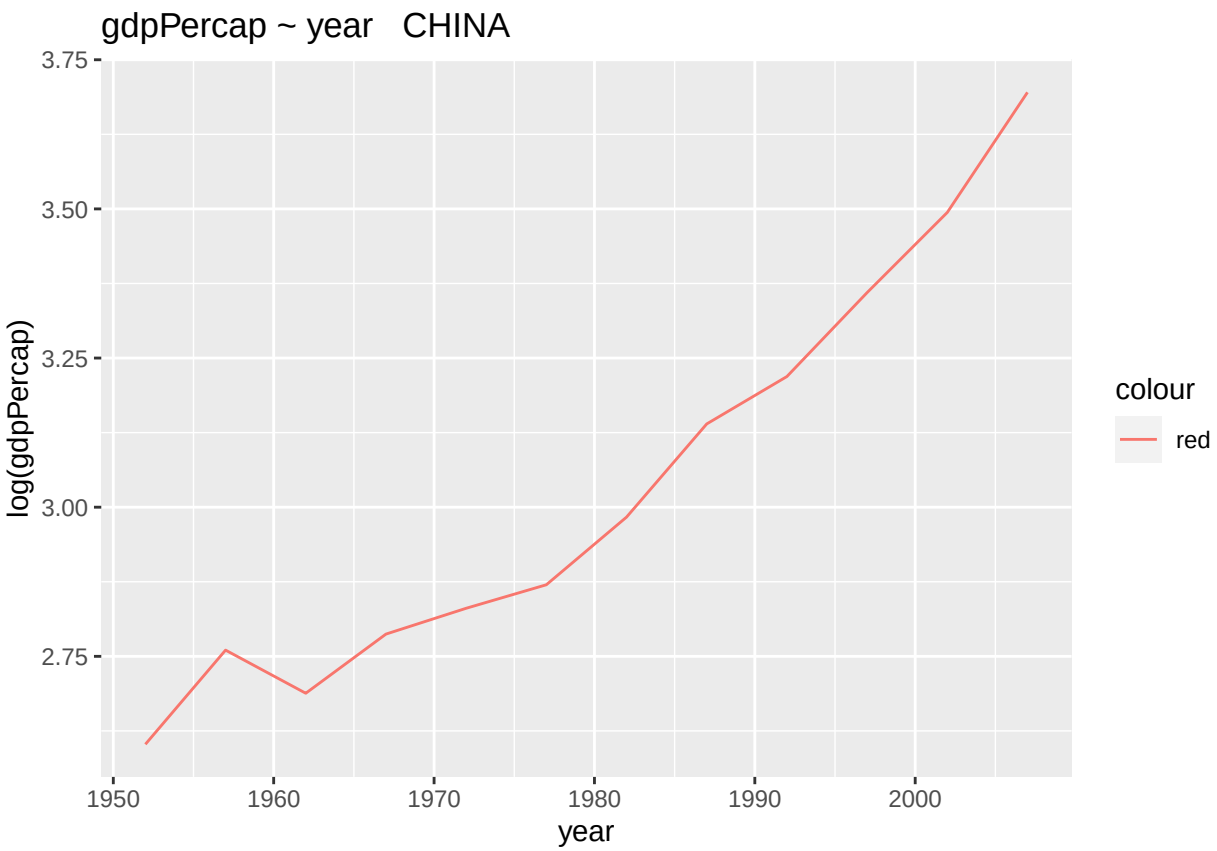


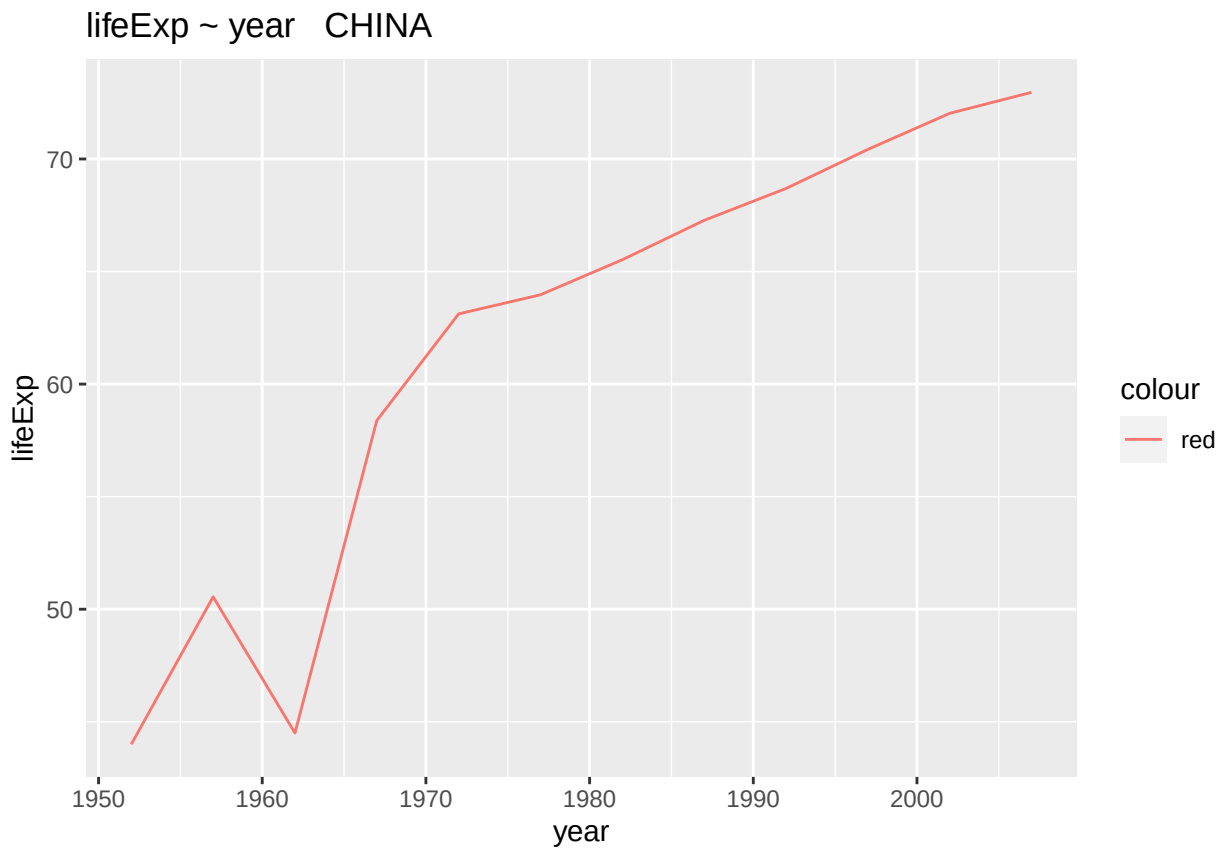
3.The relationship between lifeEXp and gdpPercap : We have the lifeEXp and gdpPercap plotted, and fit the point with linear model.According to the following plot, we can see that the lifeEXp and gdpPercap fit the linear model badly, and the fit model seems more likely to be log function.But the claim that there's a positive correlation between lifeEXp and gdpPercap.



4.A tpical example : the development of China among years We have the gdp, gpdPercap, lifeEXp plotted against year.And we can see the all round development of China.Although there is a little swing at about 1960, the macroscopic increase of gdp, gpdPercap and lifeEXp is remarkable and enormous.Besides the growing rate of these three variable is also inspiring.


gdp ~ year CHINA
12.8
12.4
12.0
11.6
log(gdp)
1950
1960
1970
1980
1990
2000
year
colour
red


gdpPercap ~ year CHINA
3.75
3.50
3.25
3.00
2.75
log(gdpPercap)
1950
1960
1970
1980
1990
2000
year
colour
red



That is all the data analysis required of gadminder.