Aging Population

Tuo

4/6/2022

# Set up

# set up librarys   
library(tidyverse)

## Warning in file(con, "r"): cannot open file '/var/db/timezone/zoneinfo/  
## +VERSION': No such file or directory

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.4 ✓ dplyr 1.0.7  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 2.0.1 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readxl)

# Load data

data <- read\_csv("data.csv", col\_names = FALSE)

## Rows: 16 Columns: 6

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (6): X1, X2, X3, X4, X5, X6

##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

View(data)  
colnames(data) <- c("year","population","out\_of\_state\_population","in\_state\_population","in\_state\_middle\_age\_population","in\_state\_aging\_population")

# Data Transformation

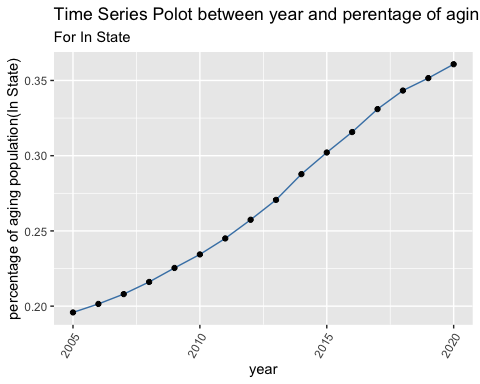
data <-  
 data %>%  
 mutate(aging\_percent\_in\_state = in\_state\_aging\_population/in\_state\_population,  
 aging\_percent\_out\_of\_state = in\_state\_aging\_population/out\_of\_state\_population,  
 aging\_percent\_total\_population = in\_state\_aging\_population/population  
 )  
head(data,5)

## # A tibble: 5 × 9  
## year population out\_of\_state\_population in\_state\_population in\_state\_middle\_…  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2005 1778. 438. 1360. 609.  
## 2 2006 1815. 467. 1368. 610.  
## 3 2007 1858. 499. 1379. 608   
## 4 2008 1888. 517. 1391. 607.  
## 5 2009 1921. 542. 1401. 602.  
## # … with 4 more variables: in\_state\_aging\_population <dbl>,  
## # aging\_percent\_in\_state <dbl>, aging\_percent\_out\_of\_state <dbl>,  
## # aging\_percent\_total\_population <dbl>

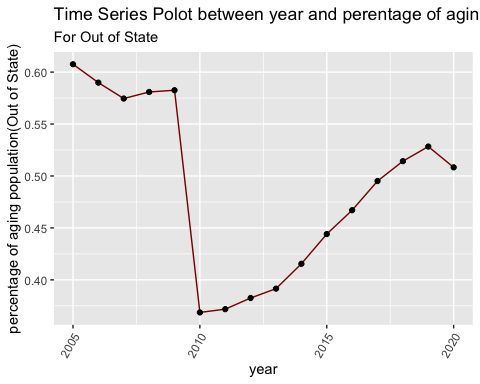
# 计算老年人口占比，参见data\_01.csv

# Data Visualization

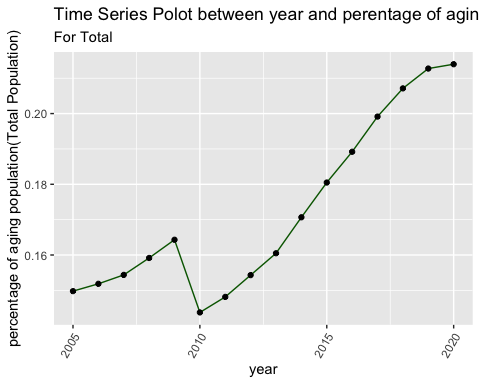
# 时间序列图，x轴-年份，y轴-户籍老年人口占户籍人口比列  
ggplot(data, aes(x=year, y=data$aging\_percent\_in\_state)) +   
 geom\_line(color="steelblue") +  
 geom\_point()+  
 xlab("") +   
 theme(axis.text.x=element\_text(angle=60, hjust=1)) +  
 labs(  
 x = "year",  
 y = "percentage of aging population(In State)",  
 title = "Time Series Polot between year and perentage of aging population",  
 subtitle = "For In State"  
 )



# 时间序列图，x轴-年份，y轴-户籍老年人口占外来人口比列  
ggplot(data, aes(x=year, y=data$aging\_percent\_out\_of\_state)) +   
 geom\_line(color="darkred") +  
 geom\_point()+  
 xlab("") +   
 theme(axis.text.x=element\_text(angle=60, hjust=1)) +   
 labs(  
 x = "year",  
 y = "percentage of aging population(Out of State)",  
 title = "Time Series Polot between year and perentage of aging population",  
 subtitle = "For Out of State"  
 )



# 时间序列图，x轴-年份，y轴-户籍老年人口占总人口比列  
ggplot(data, aes(x=year, y=data$aging\_percent\_total\_population)) +   
 geom\_line(color="darkgreen") +  
 geom\_point()+  
 xlab("") +   
 theme(axis.text.x=element\_text(angle=60, hjust=1)) +  
 labs(  
 x = "year",  
 y = "percentage of aging population(Total Population)",  
 title = "Time Series Polot between year and perentage of aging population",  
 subtitle = "For Total"  
 )



# Model building linear regression

# 线性回归模型 x-户籍人口，y-户籍老年人口占户籍人口比列  
model\_1 <- lm(aging\_percent\_in\_state ~ in\_state\_population, data = data)  
summary(model\_1)

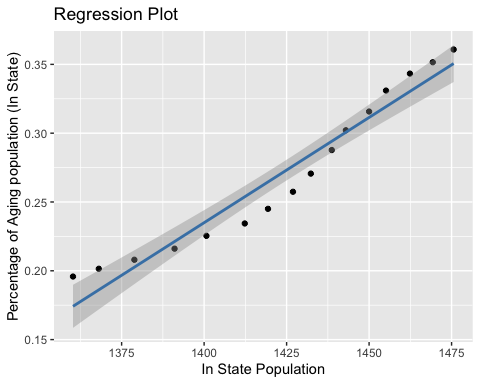
##   
## Call:  
## lm(formula = aging\_percent\_in\_state ~ in\_state\_population, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.019545 -0.011421 0.002882 0.010901 0.021649   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.9061057 0.1430706 -13.32 2.42e-09 \*\*\*  
## in\_state\_population 0.0015293 0.0001004 15.23 4.17e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01403 on 14 degrees of freedom  
## Multiple R-squared: 0.9431, Adjusted R-squared: 0.939   
## F-statistic: 231.8 on 1 and 14 DF, p-value: 4.175e-10

# 统计结果显著：p值4.175e-10，模型效果良好： R-squared:0.9431  
# 模型概括： 户籍老龄人口率 = -1.9061057 + 0.0015293 \* 户籍人口  
  
confint(model\_1)

## 2.5 % 97.5 %  
## (Intercept) -2.212961577 -1.599249804  
## in\_state\_population 0.001313901 0.001744749

# 模型95%置信区间  
  
  
ggplot(data = data, aes(x=in\_state\_population,y=aging\_percent\_in\_state))+  
 geom\_point()+  
 geom\_smooth(method=lm, color = "steelblue")+  
 labs(  
 x = "In State Population",  
 y = "Percentage of Aging population (In State)",  
 title ="Regression Plot"  
 )

## `geom\_smooth()` using formula 'y ~ x'



# 模型作图，户籍人口越多，户籍老年人口占比越大，说明老龄化加速

# Model building quadratic regression

in\_state\_population2 <- data$in\_state\_population^2  
  
model\_2 <- lm(aging\_percent\_in\_state ~ in\_state\_population + in\_state\_population2, data = data)  
summary(model\_2)

##   
## Call:  
## lm(formula = aging\_percent\_in\_state ~ in\_state\_population + in\_state\_population2,   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0111646 -0.0043089 0.0000969 0.0032301 0.0104890   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.896e+01 2.876e+00 6.592 1.74e-05 \*\*\*  
## in\_state\_population -2.792e-02 4.058e-03 -6.880 1.12e-05 \*\*\*  
## in\_state\_population2 1.038e-05 1.431e-06 7.257 6.39e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.006477 on 13 degrees of freedom  
## Multiple R-squared: 0.9887, Adjusted R-squared: 0.987   
## F-statistic: 570 on 2 and 13 DF, p-value: 2.18e-13

# 统计结果显著：p值2.18e-13，模型效果良好： R-squared:0.9887  
# 模型概括： 户籍老龄人口率 = 1.896e+01 -2.792e-02 \* 户籍人口 + 1.038e-05 \* 户籍人口^2  
  
ggplot(data = data, aes(x=in\_state\_population,y=aging\_percent\_in\_state))+  
 geom\_point()+  
 geom\_smooth(method=lm, formula = y ~ x + I(x^2), color = "red")+  
 labs(  
 title ="Quadratic Regression Plot"  
 )

