ADS_groupwork

Samuel

2024-03-13

Read the data first

Part 1: Exploring the data

Question 1

```
alcohol_data<- data_raw %>% filter(year==2019 & sex =="Male" &
        age=="40 to 44" & cause=="Alcohol use disorders" & measure=="Deaths")
head(alcohol_data)
```

```
##   measure                        location  sex      age                 cause
## 1  Deaths                  South Asia - WB Male 40 to 44 Alcohol use disorders
## 2  Deaths Middle East & North Africa - WB Male 40 to 44 Alcohol use disorders
## 3  Deaths        East Asia & Pacific - WB Male 40 to 44 Alcohol use disorders
## 4  Deaths                    North America Male 40 to 44 Alcohol use disorders
## 5  Deaths         Sub-Saharan Africa - WB Male 40 to 44 Alcohol use disorders
## 6  Deaths      Europe & Central Asia - WB Male 40 to 44 Alcohol use disorders
##    metric year         val       upper       lower
## 1 Percent 2019 0.012215856 0.014481335 0.008484016
## 2 Percent 2019 0.003040330 0.003688087 0.002506647
## 3 Percent 2019 0.012726958 0.014213882 0.008809356
## 4 Percent 2019 0.029002889 0.031514494 0.026391834
## 5 Percent 2019 0.003210615 0.004246450 0.002634772
## 6 Percent 2019 0.053798538 0.058466137 0.047957598
```

```
# comparing
highest_rate<-sort(alcohol_data$val,decreasing = TRUE)[1]
alcohol_data %>% filter(val == highest_rate) %>% select(location) #result
```

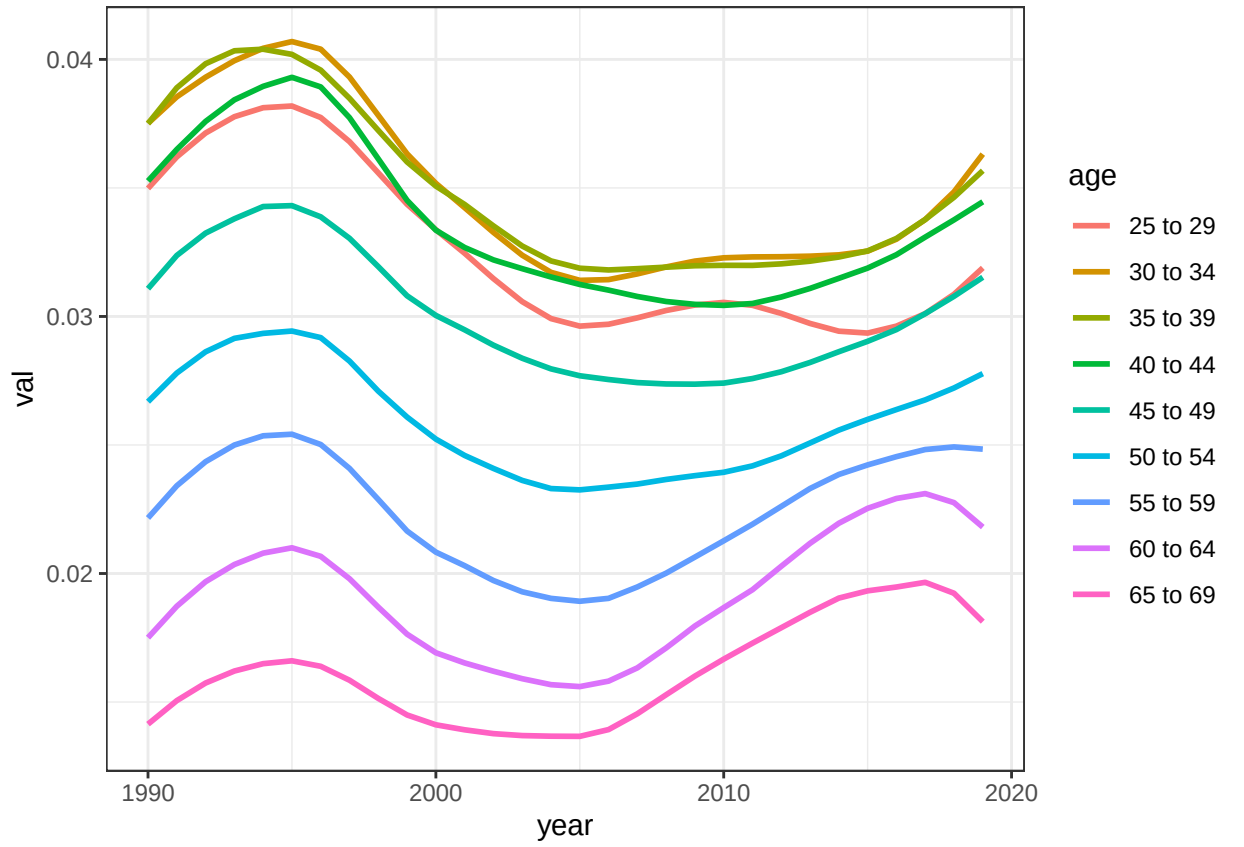
```
##                     location
## 1 Europe & Central Asia - WB
```

It show that Europe & Central Asia is the region of the world has the highest rate of alcohol-related deaths among men aged 40-44.

Question 2

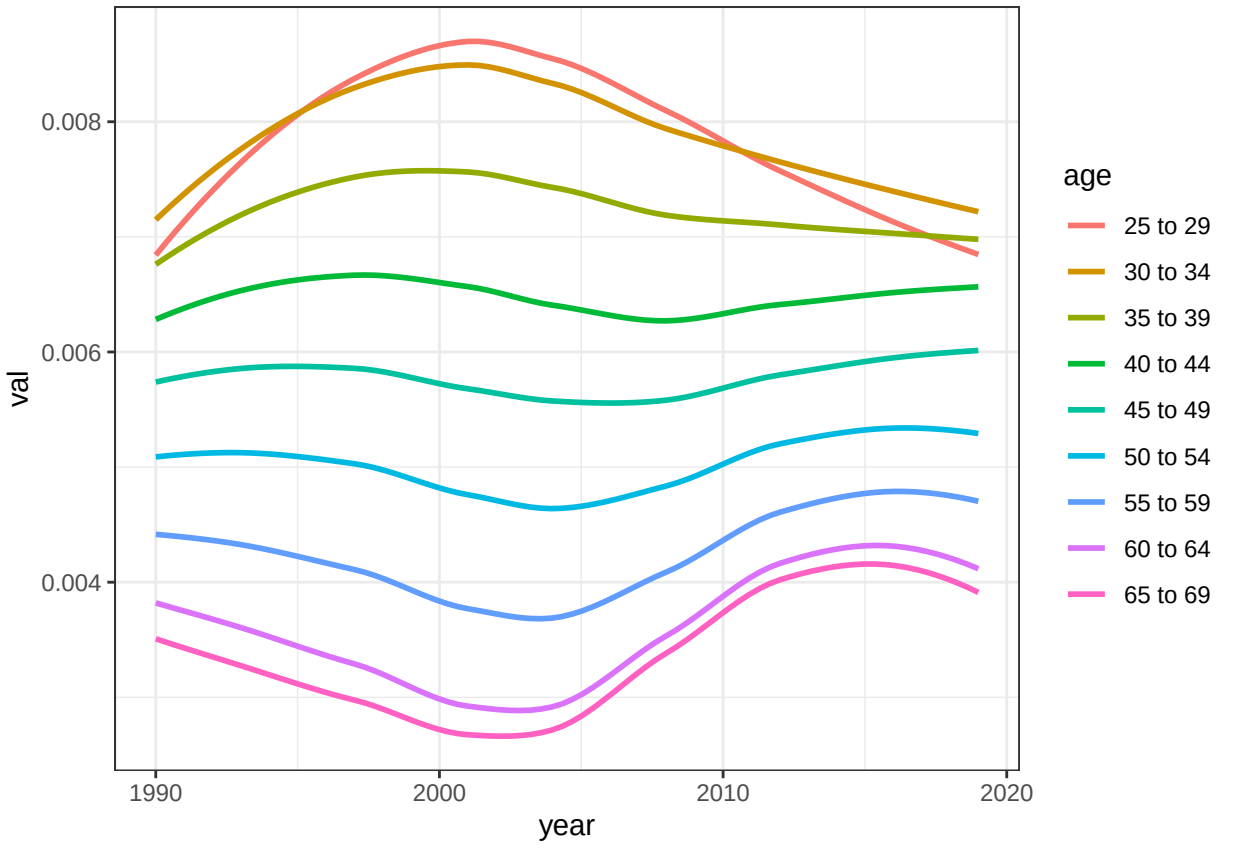
```
library(ggplot2)
alcoholdisease_man<- data_raw %>% filter(location =="East Asia & Pacific - WB" &
         cause=="Alcohol use disorders" & measure=="Prevalence" & sex== "Male")

alcoholdisease_woman<- data_raw %>% filter(location =="East Asia & Pacific - WB"
 & cause=="Alcohol use disorders" & measure=="Prevalence" & sex== "Female")
#plot it
ggplot(data = alcoholdisease_man, aes(x=year,y=val,color=age)) +
  geom_line(linewidth=1) +theme_bw()
```



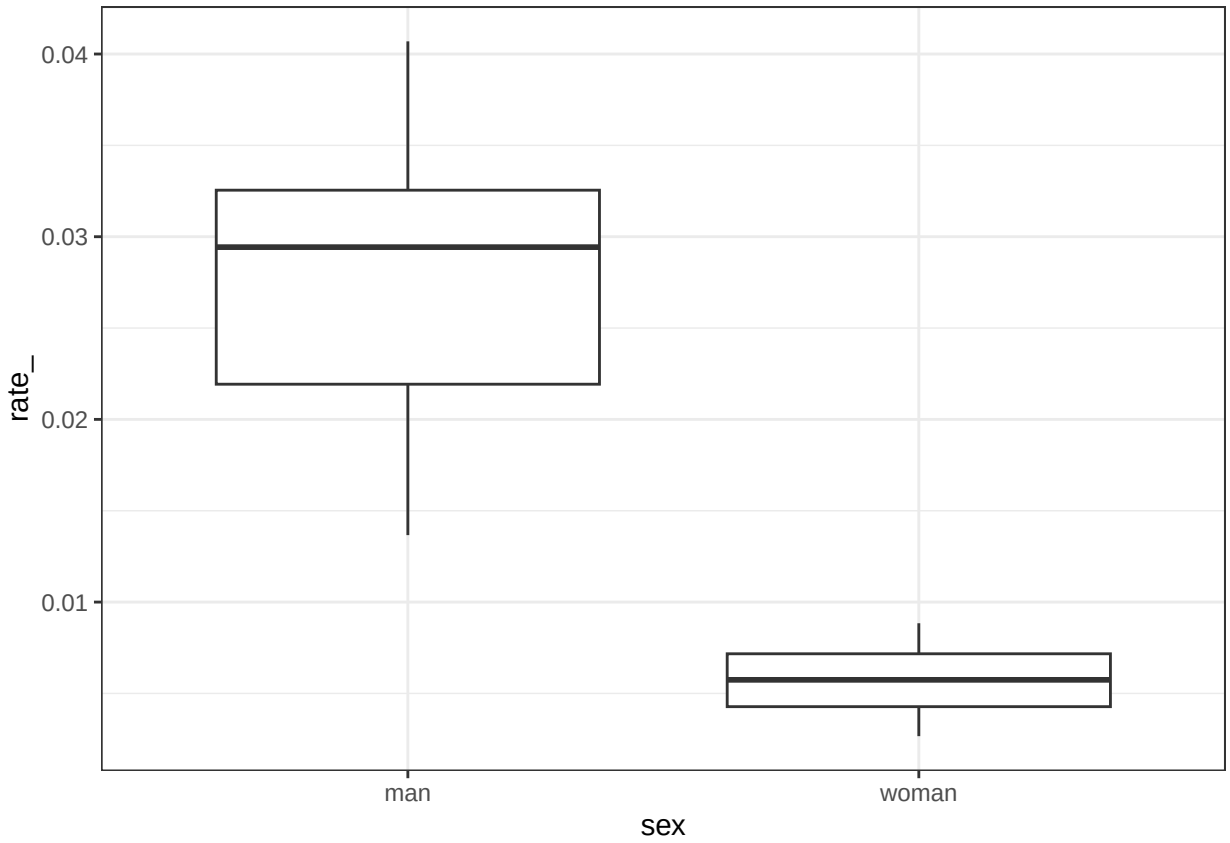

```
ggplot(data = alcoholdisease_woman, aes(x=year,y=val,color=age)) +
  geom_smooth(method = "auto", se = FALSE) +theme_bw() #2 method choose which
```

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




```
#comparing between male and female
man_rate<-alcoholdisease_man %>% select(val)
woman_rate<-alcoholdisease_woman %>% select(val)
result_comparing<-cbind(man_rate,woman_rate)
colnames(result_comparing)= c("man","woman")
library(reshape2)
result_comparing = melt(result_comparing,id.vars = c())
colnames(result_comparing)= c("sex","rate_")

result_comparing$sex = as.factor(result_comparing$sex)
ggplot(data = result_comparing,aes(x=sex,y=rate_))+geom_boxplot()+theme_bw()
```



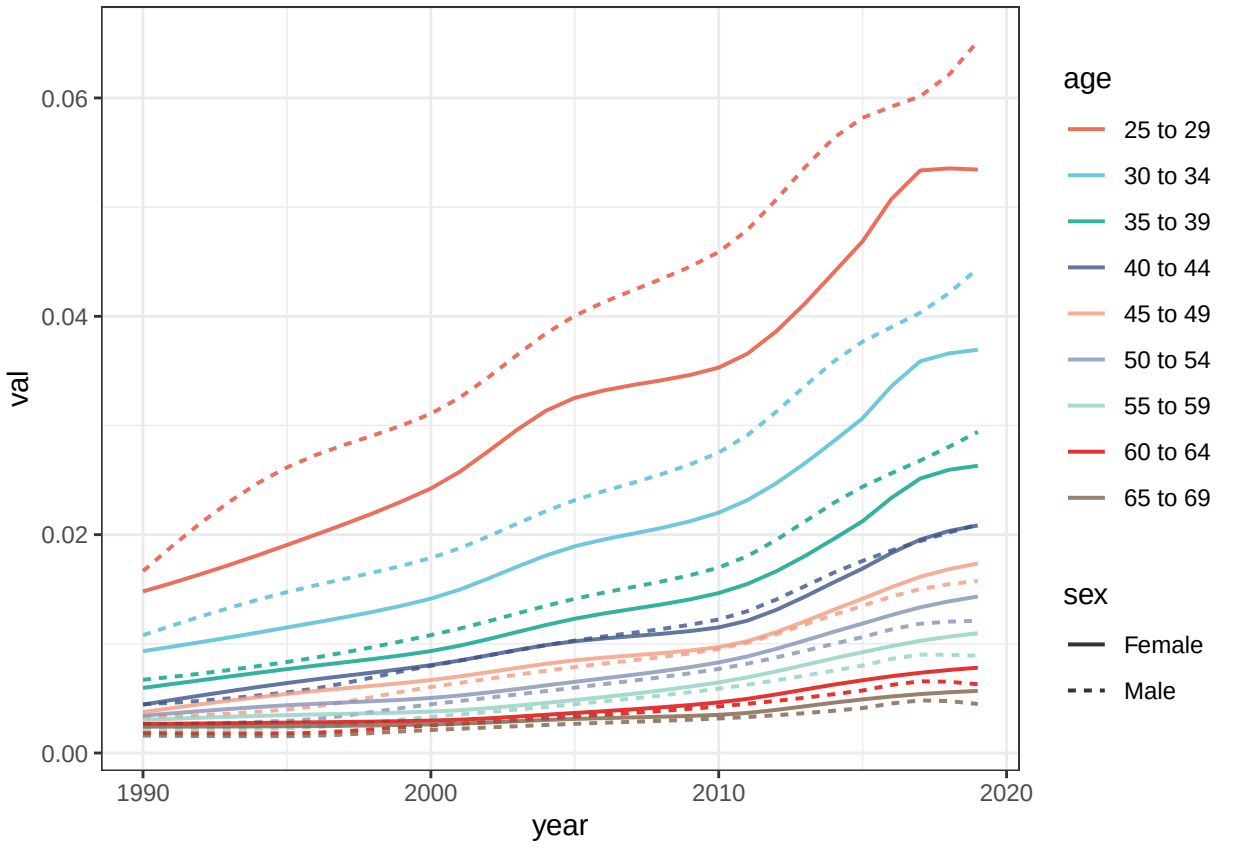
There is a significant difference between men and women at the prevalence of alcohol-related disease in the East Asia and Pacific region

Question 3

There is not specific "the United States" in the column "location". So the most related region "North America" is chosen.

```
library(ggsci)
#
opioid_use<- data_raw %>% filter(location =="North America"
 & cause=="Opioid use disorders" & measure=="Prevalence")

ggplot(data = opioid_use, aes(x=year,y=val,color=age,linetype =sex)) +
  geom_line(linewidth=0.7,alpha =0.8) +theme_bw()+scale_color_npg()
```




```
result_cor_num<-c()
age_group<-unique(opioid_use$age)

for (i in age_group){
  for(j in c("Male","Female")){
opioid_use_<-opioid_use %>% filter(sex==j & age==i)
result_cor_num<-append(result_cor_num,cor(opioid_use_$year,opioid_use_$val,method="pearson",use="comple


result_cor<-data.frame(Sex=rep(c("Male","Female"),length(age_group)),Age=rep(age_group,each=2),cor_num=
result_cor
```

```
##       Sex      Age   cor_num
## 1    Male 25 to 29 0.9949676
## 2  Female 25 to 29 0.9835978
## 3    Male 30 to 34 0.9793306
## 4  Female 30 to 34 0.9664326
## 5    Male 35 to 39 0.9705637
## 6  Female 35 to 39 0.9553712
## 7    Male 40 to 44 0.9768722
## 8  Female 40 to 44 0.9609815
## 9    Male 45 to 49 0.9806583
## 10 Female 45 to 49 0.9627802
## 11   Male 50 to 54 0.9806275
## 12 Female 50 to 54 0.9567769
```

```
## 13   Male 55 to 59 0.9770511
## 14 Female 55 to 59 0.9441719
## 15   Male 60 to 64 0.9724931
## 16 Female 60 to 64 0.9311951
## 17   Male 65 to 69 0.9722509
## 18 Female 65 to 69 0.9265331
```

From the figure we can see in both Female and Male group age 25-29 is most affected