Part2

Samuel

2024-04-10

Question:

We often hear that middle age(35 to 59) males are more likely to hit the bottle due to encountering setback in their careers or losing their beloveds and they are more likely to have sudden death because the disorderly usage of alcohol. So can we prove that the middle age males are more likely to die because the disorderly usage of alcohol?

Answer

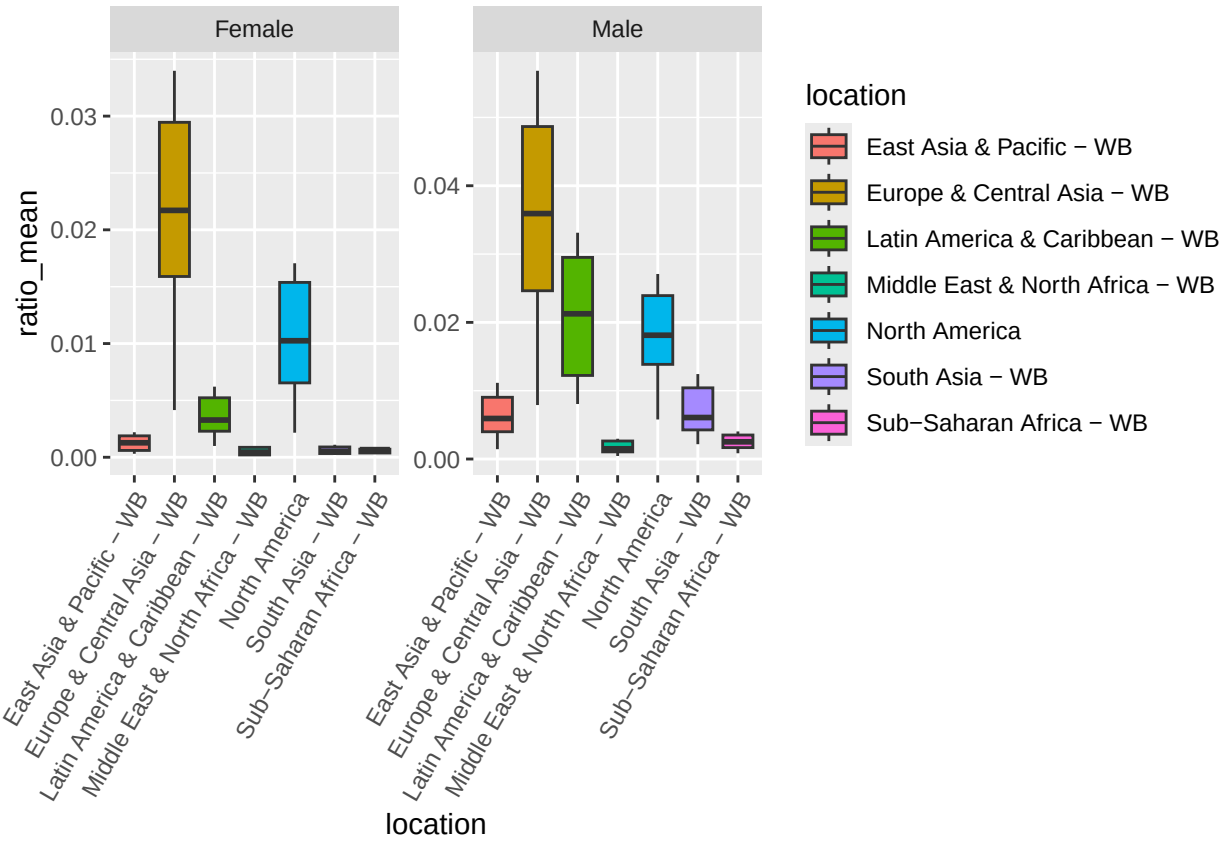
First we should take a brief overview of the whole data.

```
library(dplyr)
```

```
##
##     'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
data_raw<-read.csv("C:/Users/Samuel/Downloads/substance_use.csv")
locations<-unique(data_raw$location)
ages<-unique(data_raw$age)
sexs<-c("Male","Female")
years<-c(1990:2019)
ratio<-c()
ratio_mean<-c()
ratio_wo<-c()
for (i in locations){
    Alcohol_deaths<- data_raw %>% filter(location ==i
                    & cause=="Alcohol use disorders" & measure=="Deaths")
    Alcohol_deaths<-Alcohol_deaths[order(Alcohol_deaths$age),]$val #sorting by year
  for (j in c(1:length(ages))){
    ratio<-c()
```

```
    ratio_wo<-c()
    for(k in c(0:length(years)-1)){
      s<-(j-1)*60+1
      ratio<-append(ratio,(Alcohol_deaths[s+k*2]))
      ratio_wo<-append(ratio_wo,(Alcohol_deaths[s+k*2+1]))
}
    ratio_mean<-append(ratio_mean,mean(ratio))
    ratio_mean<-append(ratio_mean,mean(ratio_wo))}}


result_diff_AD<-data.frame(age=rep(ages,time = length(locations)*2),
      location=rep(locations,each=length(ages)*2),
      sex=rep(sexs,time = length(locations)*length(ages)),ratio_mean)

ggplot(data=result_diff_AD,aes(x=location,y=ratio_mean,fill=location))+
  geom_boxplot()+facet_wrap(~sex,scales = "free_y")+
  theme(axis.text.x = element_text(angle = 60,hjust = 1))
```

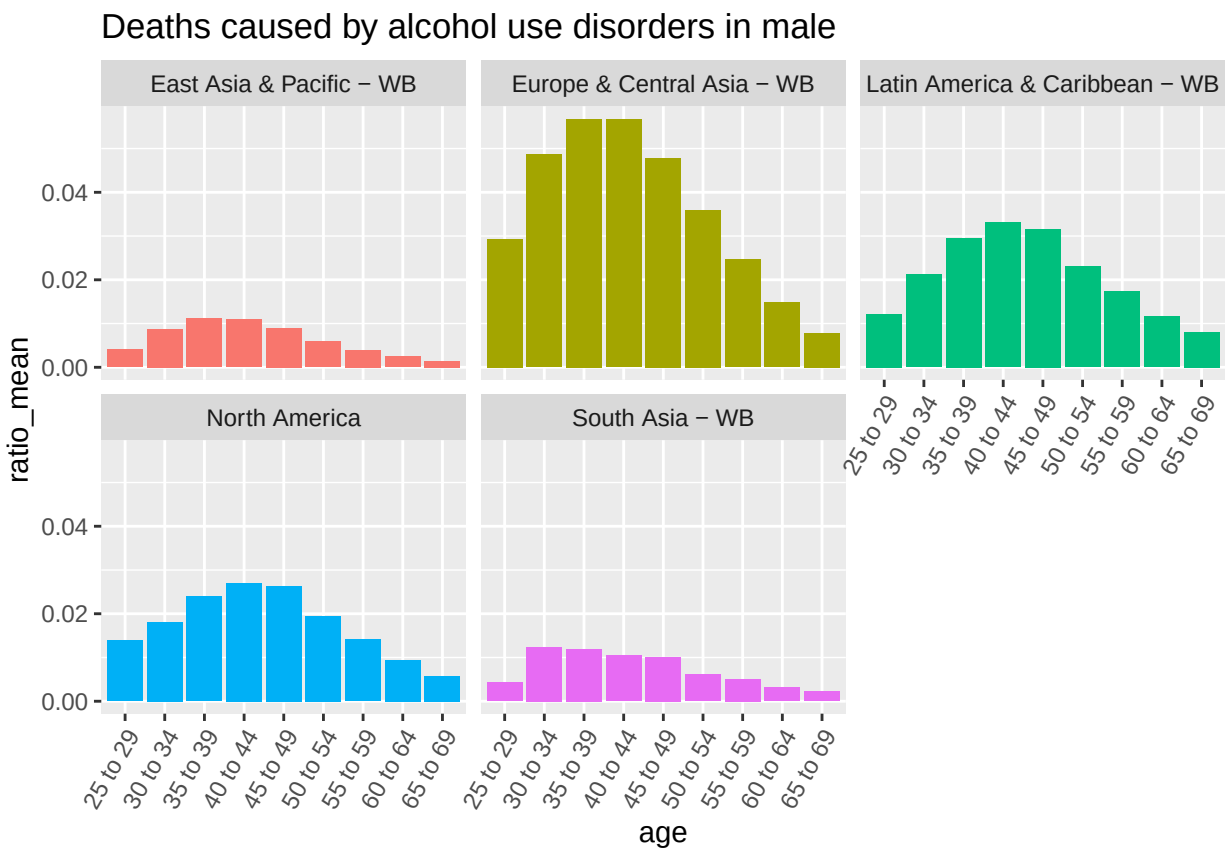


From the bar plot we get, we can see that Europe & Central Asia has highest average level of rate. Besides, we can also find that the level of rate from Sub-Saharan Africa and Middle East & North Africa is significantly lower than the other part of the locations in both "Male" and "Female" groups so we can take them as out liners and drop them out.

– Both Sub-Saharan Africa and Middle East & North Africa are mostly "underdeveloped areas". People in those places have fewer chances to have a heavy alcohol exposure.

Then we want to find the differences of deaths caused by alcohol use disorders between male in different age.

```
ratio<-c()
ratio_mean<-c()
locations<-c("East Asia & Pacific - WB","North America","South Asia - WB",
"Latin America & Caribbean - WB","Europe & Central Asia - WB")
for (i in locations){
      Alcohol_deaths<- data_raw %>% filter(location ==i
                    & cause=="Alcohol use disorders" & measure=="Deaths")
    Alcohol_deaths<-Alcohol_deaths[order(Alcohol_deaths$age),]$val #sorting by year
  for (j in c(1:length(ages))){
    ratio<-c()
    for(k in c(0:length(years)-1)){
      s<-(j-1)*60+1
      ratio<-append(ratio,(Alcohol_deaths[s+k*2]))
}
    ratio_mean<-append(ratio_mean,mean(ratio))}
    }

result_diff_A_D_inMale<-data.frame(age=rep(ages,length(locations)),
    location=rep(locations,each=length(ages)),ratio_mean)
ggplot(data=result_diff_A_D_inMale,aes(x=age,y=ratio_mean,fill=location))+
geom_bar(stat="identity")+facet_wrap(~location)+
theme(legend.position = "none",axis.text.x = element_text(angle = 60,hjust = 1))+
ggtitle("Deaths caused by alcohol use disorders in male")
```

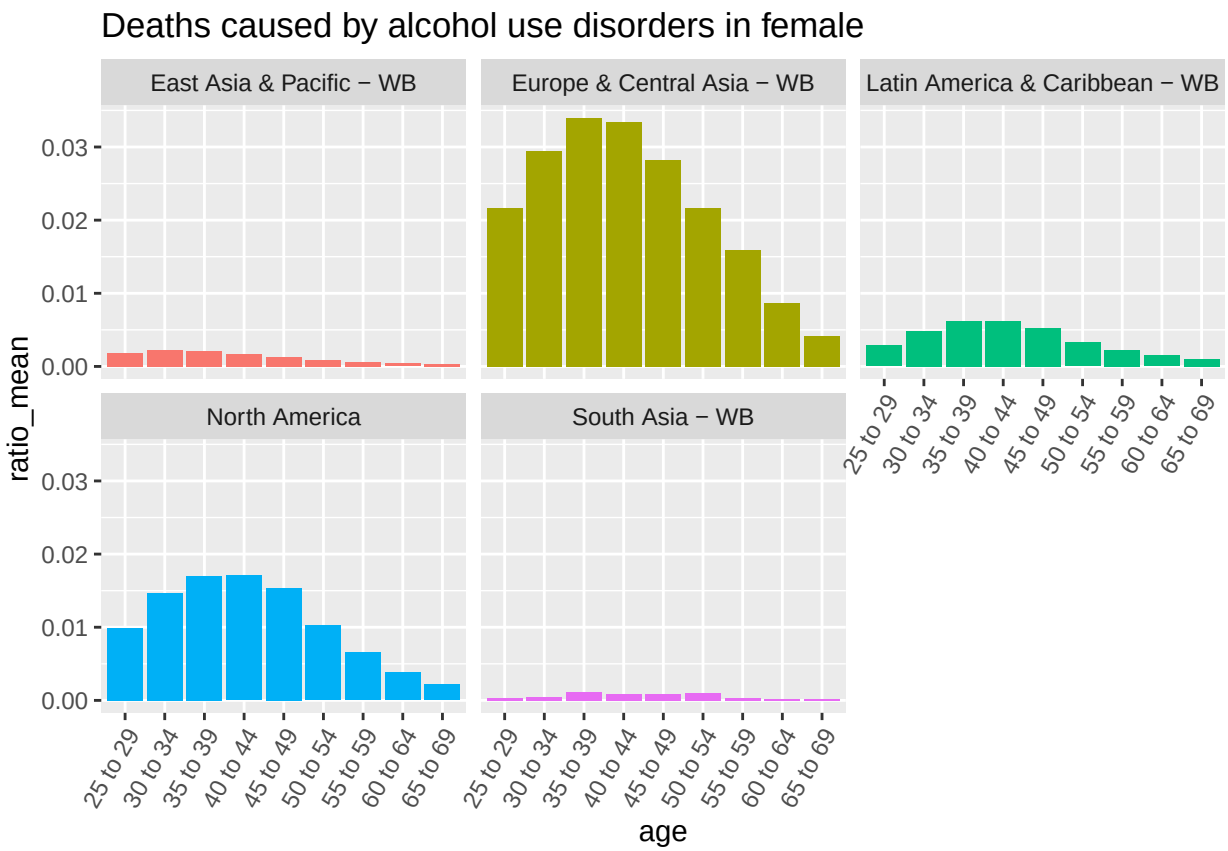


It shows that all groups have significant difference in total.

In a similar way, we want to find the differences of deaths caused by alcohol use disorders between female in

different age.

```
ratio<-c()
ratio_mean<-c()
for (i in locations){
      Alcohol_deaths<- data_raw %>% filter(location ==i
                    & cause=="Alcohol use disorders" & measure=="Deaths")
    Alcohol_deaths<-Alcohol_deaths[order(Alcohol_deaths$age),]$val #sorting by year
  for (j in c(1:length(ages))){
    ratio<-c()
    for(k in c(0:length(years)-1)){
      s<-(j-1)*60+1
      ratio<-append(ratio,(Alcohol_deaths[s+k*2+1]))
}
    ratio_mean<-append(ratio_mean,mean(ratio))}
    }

result_diff_AD<-data.frame(age=rep(ages,length(locations)),location=rep(locations,each=length(ages)),ra
ggplot(data=result_diff_AD,aes(x=age,y=ratio_mean,fill=location))+
geom_bar(stat="identity")+facet_wrap(~location)+
theme(legend.position = "none",axis.text.x = element_text(angle = 60,hjust = 1))+
ggtitle("Deaths caused by alcohol use disorders in female")
```



It shows that there is only 2 groups with a significant difference in total.

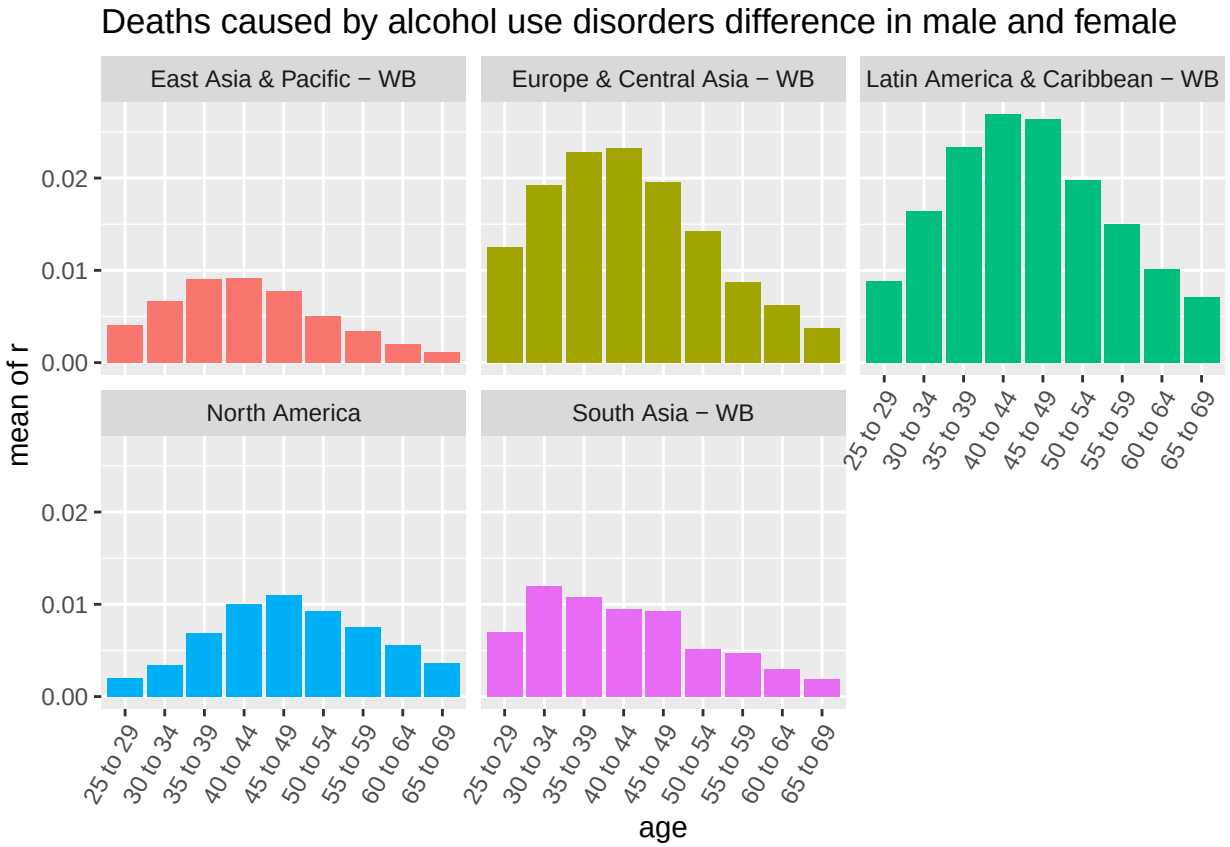
Then,to ensure that the rise of the deaths caused by alcohol use disorders affect more apparently on male than female in the same age. We take the female data as a reference value and define the difference of Deaths

rate contributing to alcohol use disorders in a year between female and male as $r = Male\,rate\,of\,deaths - Female\,rate\,of\,deaths$ and we take the mean of each year(during 1990 to 2019) for each age group and get $\bar{r}$.

– We choose $Male\,rate\,of\,deaths - Female\,rate\,of\,deaths$ instead of $\frac{Male\,rate\,of\,deaths}{Female\,rate\,of\,deaths}$ to avoid the minimal numbers pairs(such as a value in 0.0001 and another in 0.0002) badly interfere the result.

If $\bar{r}$ shows significantly highest in the middle ages, it means male are more likely to encounter deaths caused by alcohol use disorders than female, especially in the middle age.

```
years<-c(1990:2019)
ratio<-c()
ratio_mean<-c()
for (i in locations){
      Alcohol_deaths<- data_raw %>% filter(location ==i
                      & cause=="Alcohol use disorders" & measure=="Deaths")
    Alcohol_deaths<-Alcohol_deaths[order(Alcohol_deaths$age),]$val #sorting by year
  for (j in c(1:length(ages))){
    ratio<-c()
    for(k in c(0:length(years)-1)){
      s<-(j-1)*60+1
      ratio<-append(ratio,(Alcohol_deaths[s+k*2]-Alcohol_deaths[s+k*2+1]))
}
    ratio_mean<-append(ratio_mean,mean(ratio))}
    }

result_diff_AD<-data.frame(age=rep(ages,length(locations)),
                       location=rep(locations,each=length(ages)),ratio_mean)
ggplot(data=result_diff_AD,aes(x=age,y=ratio_mean,fill=location))+
geom_bar(stat="identity")+facet_wrap(~location)+
theme(legend.position = "none",axis.text.x = element_text(angle = 60,hjust = 1))+
ggtitle("Deaths caused by alcohol use disorders difference in male and female")+
ylab("mean of r")
```



It shows that comparing with "female group effect", deaths caused by alcohol use disorders affect more apparently on male and this effect in male enrich in the "middle age group" (35 to 59). In conclusion, middle age males are more likely to encounter deaths causing by the disorderly usage of alcohol.