Yufei_HW4

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#Problem 1
# Download the two complete datasets
library(readr)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyvers
e 1.2.1 ──
```

```
## ✔ ggplot2 3.2.1     ✔ purrr   0.3.3
## ✔ tibble  2.1.3     ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0     ✔ stringr 1.4.0
## ✔ ggplot2 3.2.1     ✔ forcats 0.4.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conf
licts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(tidyr)

#The Domestic general government health expenditure dataset
dataset1 <- read_csv("~/Desktop/GHED_GGHE-D_pc_PPP_SHA2011(1).csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Warning: Duplicated column names deduplicated: 'Domestic general
government
## health expenditure (GGHE-D) per capita in PPP int$' => 'Domestic
general
## government health expenditure (GGHE-D) per capita in PPP int$_1'
[3],
## 'Domestic general government health expenditure (GGHE-D) per capi
ta in
## PPP int$' => 'Domestic general government health expenditure (GGH
E-D) per
## capita in PPP int$_2' [4], 'Domestic general government health ex
penditure
## (GGHE-D) per capita in PPP int$' => 'Domestic general government
health
## expenditure (GGHE-D) per capita in PPP int$_3' [5], 'Domestic gen
eral
## government health expenditure (GGHE-D) per capita in PPP int$' =>
## 'Domestic general government health expenditure (GGHE-D) per capi
ta in
## PPP int$_4' [6], 'Domestic general government health expenditure
(GGHE-D)
## per capita in PPP int$' => 'Domestic general government health ex
penditure
## (GGHE-D) per capita in PPP int$_5' [7], 'Domestic general governm
ent
## health expenditure (GGHE-D) per capita in PPP int$' => 'Domestic
general
## government health expenditure (GGHE-D) per capita in PPP int$_6'
[8],
## 'Domestic general government health expenditure (GGHE-D) per capi
ta in
## PPP int$' => 'Domestic general government health expenditure (GGH
E-D) per
## capita in PPP int$_7' [9], 'Domestic general government health ex
penditure
## (GGHE-D) per capita in PPP int$' => 'Domestic general government
health
## expenditure (GGHE-D) per capita in PPP int$_8' [10], 'Domestic ge
neral
```

```
## government health expenditure (GGHE-D) per capita in PPP int$' =>
## 'Domestic general government health expenditure (GGHE-D) per capi
ta in
## PPP int$_9' [11], 'Domestic general government health expenditure
(GGHE-D)
## per capita in PPP int$' => 'Domestic general government health ex
penditure
## (GGHE-D) per capita in PPP int$_10' [12], 'Domestic general gover
nment
## health expenditure (GGHE-D) per capita in PPP int$' => 'Domestic
general
## government health expenditure (GGHE-D) per capita in PPP int$_11'
[13],
## 'Domestic general government health expenditure (GGHE-D) per capi
ta
## in PPP int$' => 'Domestic general government health expenditure (
GGHE-
## D) per capita in PPP int$_12' [14], 'Domestic general government
health
## expenditure (GGHE-D) per capita in PPP int$' => 'Domestic general
## government health expenditure (GGHE-D) per capita in PPP int$_13'
[15],
## 'Domestic general government health expenditure (GGHE-D) per capi
ta
## in PPP int$' => 'Domestic general government health expenditure (
GGHE-
## D) per capita in PPP int$_14' [16], 'Domestic general government
health
## expenditure (GGHE-D) per capita in PPP int$' => 'Domestic general
## government health expenditure (GGHE-D) per capita in PPP int$_15'
[17],
## 'Domestic general government health expenditure (GGHE-D) per capi
ta
## in PPP int$' => 'Domestic general government health expenditure (
GGHE-
## D) per capita in PPP int$_16' [18], 'Domestic general government
health
## expenditure (GGHE-D) per capita in PPP int$' => 'Domestic general
## government health expenditure (GGHE-D) per capita in PPP int$_17'
[19]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_1` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_2` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_3` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_4` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_5` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_6` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_7` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_8` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_9` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_10` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_11` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_12` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_13` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_14` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_15` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_16` = col_double(),
##   `Domestic general government health expenditure (GGHE-D) per ca
pita in PPP int$_17` = col_double()
## )
```

```
#The Adult mortality rate dataset
dataset2<- read_csv("~/Desktop/WHOSIS_000004(2).csv")
```

```
## Warning: Missing column names filled in: 'X1' [1], 'X2' [2]
```

```
## Warning: Duplicated column names deduplicated: 'Adult mortality r
ate
## (probability of dying between 15 and 60 years per 1000 population
)' =>
## 'Adult mortality rate (probability of dying between 15 and 60 yea
rs per
## 1000 population)_1' [4], 'Adult mortality rate (probability of dy
ing
## between 15 and 60 years per 1000 population)' => 'Adult mortality
rate
## (probability of dying between 15 and 60 years per 1000 population
)_2' [5]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   `Adult mortality rate (probability of dying between 15 and 60 y
ears per 1000 population)` = col_character(),
##   `Adult mortality rate (probability of dying between 15 and 60 y
ears per 1000 population)_1` = col_character(),
##   `Adult mortality rate (probability of dying between 15 and 60 y
ears per 1000 population)_2` = col_character()
## )
```

```
#Problem 1
library(dplyr)
#Make the first table tidy
names(dataset1) <- dataset1[1,]
```

```
## Warning: Must use a character vector as names.
## This warning is displayed once per session.
```

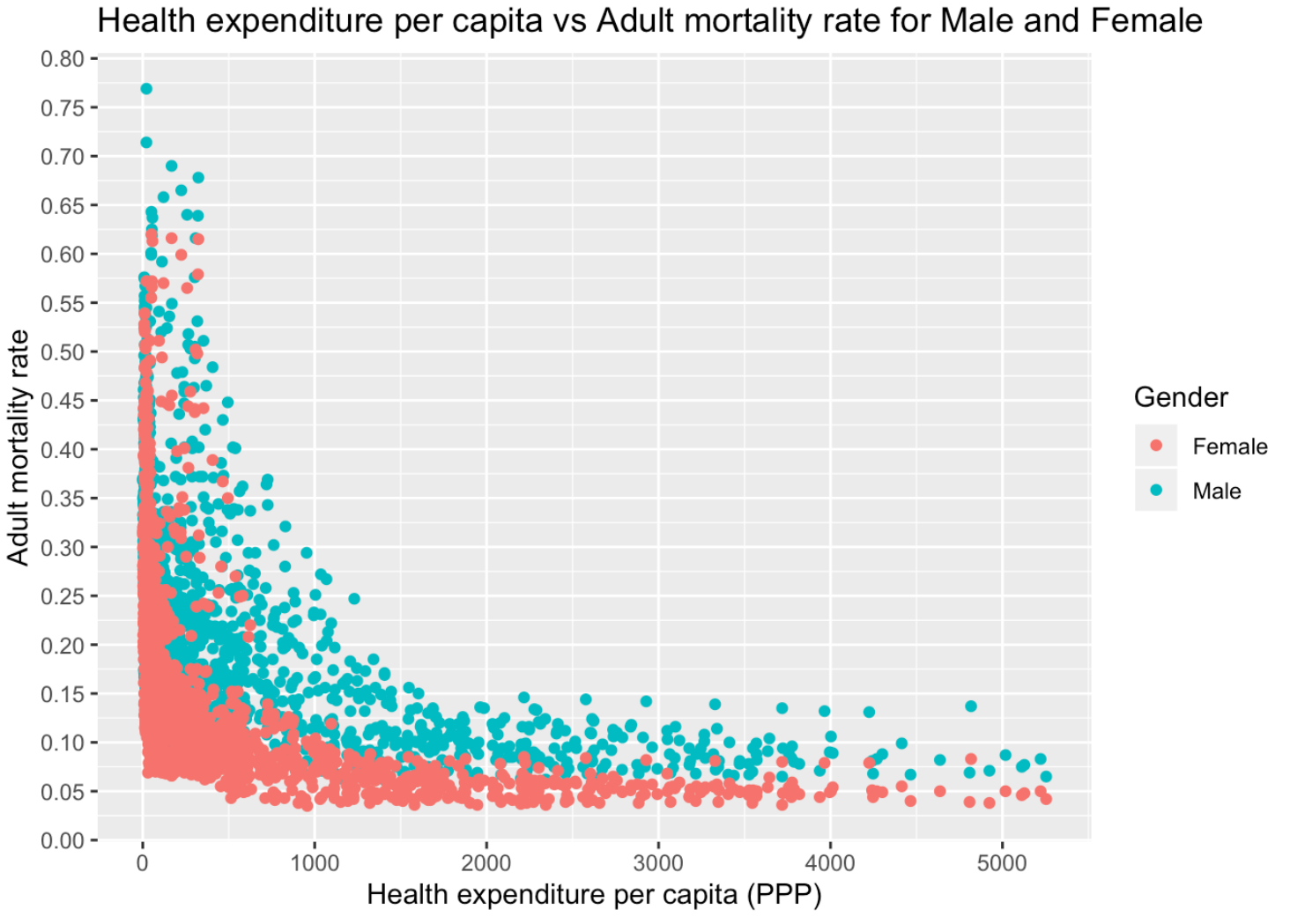
```
datset11.filtered <- dataset1[2:nrow(dataset1), ]
newdataset1 <-datset11.filtered %>% gather("2017","2016","2015","201
4","2013","2012","2011","2010","2009","2008","2007","2006","2005","2
004","2003","2002","2001","2000",key="Year",value="Values")
newdataset1$Year = as.numeric(as.character(newdataset1$Year))


#Make the second table tidy
names(dataset2) <- dataset2[1,]
datset22.filtered <- dataset2[2:nrow(dataset2), ]
newdataset2 <- datset22.filtered %>% gather("Both sexes","Male","Fem
ale",key="Gender",value="Mortality_Rate_Per_1000_Population")
newdataset2$Year = as.numeric(as.character(newdataset2$Year))
newdataset2$Mortality_Rate_Per_1000_Population = as.numeric(as.chara
cter(newdataset2$Mortality_Rate_Per_1000_Population))

# join the two tables and use head() to present the new data table
newdataset3 <- left_join(x=newdataset2,y=newdataset1,by=c("Country",
"Year"))
head(newdataset3)
```

```
## # A tibble: 6 x 5
##   Country      Year Gender     Mortality_Rate_Per_1000_Population
Values
##   <chr>       <dbl> <chr>                                      <dbl>
<dbl>
## 1 Afghanistan  2016 Both sexes                                   245
8.9
## 2 Afghanistan  2015 Both sexes                                   233
9.5
## 3 Afghanistan  2014 Both sexes                                   234
8.8
## 4 Afghanistan  2013 Both sexes                                   235
8.3
## 5 Afghanistan  2012 Both sexes                                   242
6.3
## 6 Afghanistan  2011 Both sexes                                   248
7.8
```

```
#Problem 2
 newdataset3 %>% filter(Gender==c("Male","Female")) %>% ggplot()+
  geom_point(mapping=aes(x=Values,y=(Mortality_Rate_Per_1000_Populat
ion/1000),color=Gender))+
  scale_x_continuous(name="Health expenditure per capita (PPP)") +
  scale_fill_discrete() +
  scale_y_continuous(name="Adult mortality rate", breaks=seq(0,1,0.0
5)) +
  ggtitle("Health expenditure per capita vs Adult mortality rate for
Male and Female")
```

```
## Warning in Gender == c("Male", "Female"): longer object length is
not a
## multiple of shorter object length
```

```
## Warning: Removed 100 rows containing missing values (geom_point).
```



Answer: As you can see the plot above, for both male and female, the adult mortality rate decreases when the total health expenditure per capita increases. Also, the overall adult mortality rate of female is less than the overall adult mortality rate of male.The gap of life expectancy between women and men helps to explain my observation, because this survey indicated that the global average life expectancy increased by 5.5 years between 2000 and 2016,and women live longer than men all around the world.

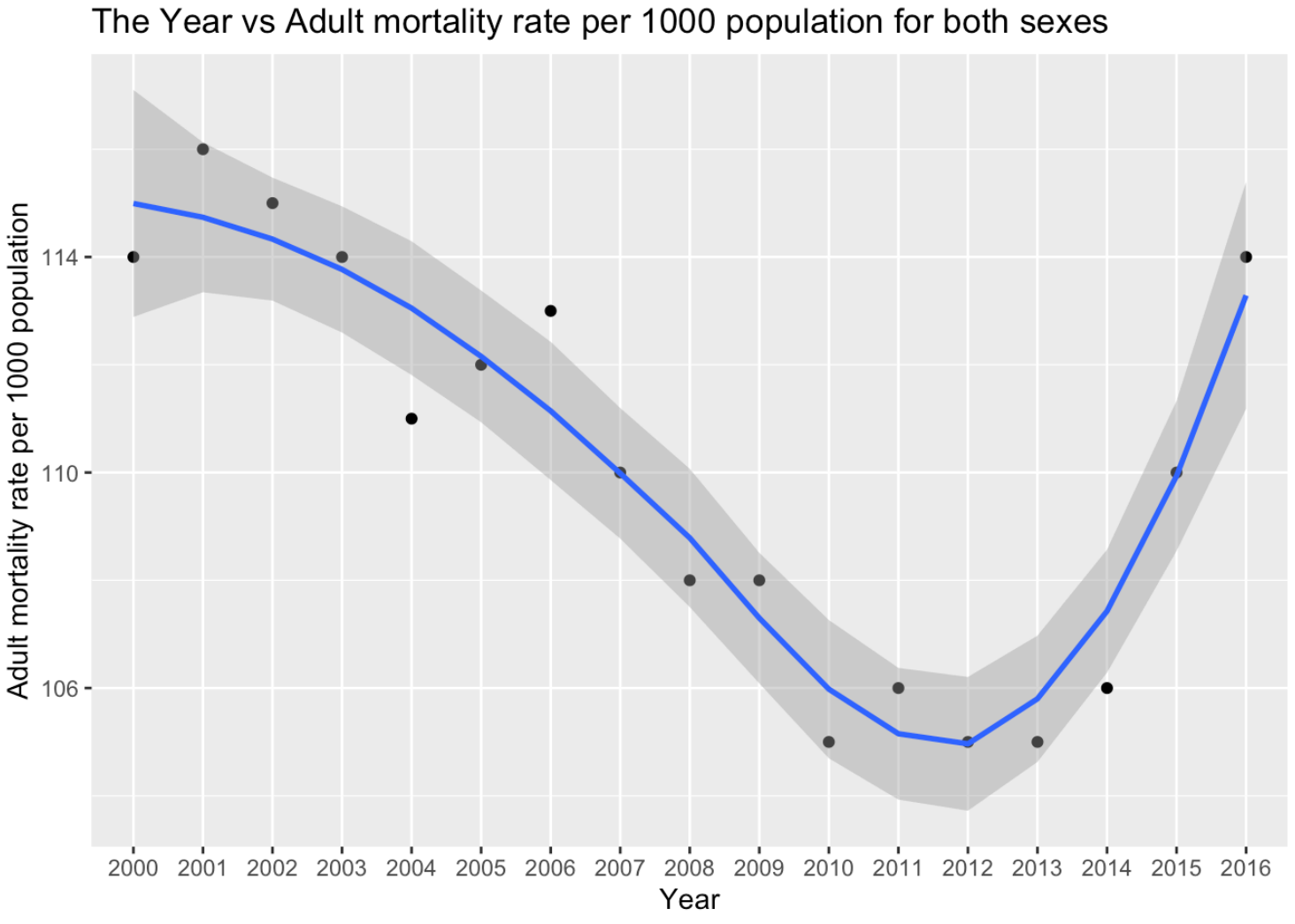
```
#Problem 3
newdataset4 <- newdataset3 %>% filter(Gender=="Both sexes")
newdataset4$Year <- as.character(newdataset4$Year )

  ggplot(data=newdataset4, aes(x=Year,y=Mortality_Rate_Per_1000_Popu
lation))+
  geom_boxplot()+
  geom_smooth(method = "auto", aes(group=1))+
scale_fill_discrete() +
scale_y_continuous(name="Adult mortality rate per 1000 population")
+
ggtitle("The Year vs Adult mortality rate per 1000 population for bo
th sexes ")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "
cs")'
```

Answer: The smooth line shows that for both sexes, the aduly mortality rate per 1000 population for both sexes decreases between 2000 and 2016. As you can see from this chart, the median adult mortality is around 0.17, the maximum adult mortality rate is around 0.7. Most outliers are in the 4th quartile group with large ranges, which means that there are obviously various adult mortality rates for both sexes in different countries.

```
#Problem 4
newdataset5 <- newdataset3 %>% filter(Gender=="Both sexes",Country==
"United States of America")
newdataset5$Year <- as.character(newdataset5$Year )

  ggplot(data=newdataset5,aes(x=Year,y=Mortality_Rate_Per_1000_Popul
ation)) +
  geom_point()+
  geom_smooth(method = "auto", aes(group=1))+
scale_fill_discrete() +
scale_y_continuous(name="Adult mortality rate per 1000 population")
+
  xlab("Year")+
ggtitle("The Year vs Adult mortality rate per 1000 population for bo
th sexes ")
```

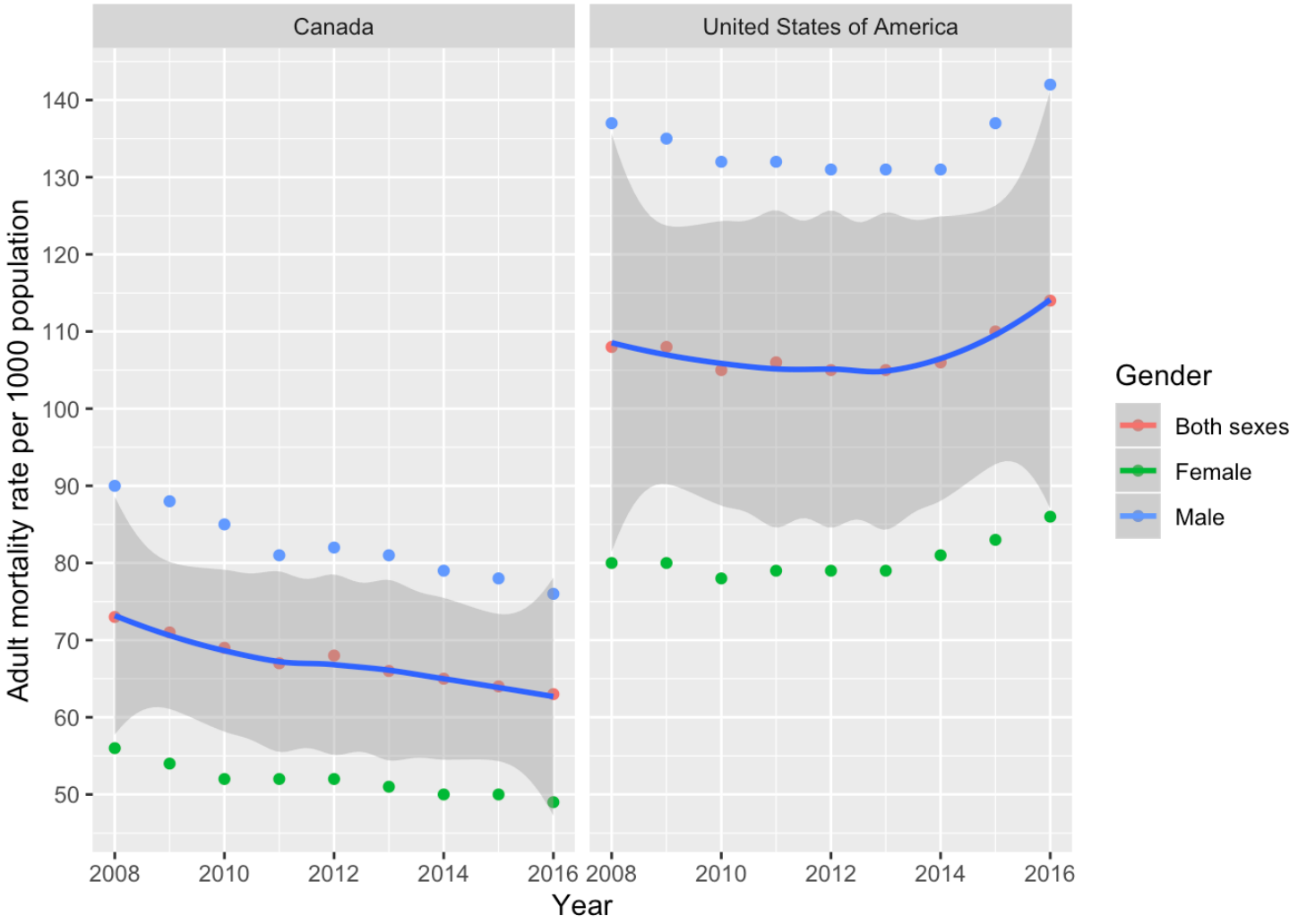
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Answer: It matches with problem 2 that the mortality rates for male and female decrease with the increase of the health expenditure from 2000 to 2012. However, from the drug overdose death rates table, the death rates increase drmatically from 2000 to 2010 for all persons, and the birth rating decreases at the same time. Therefore, there is a concave-up shape in the grapgh,and the adult mortality rates in US increases after 2012.

```
#Problem 5
newdataset3 %>% filter(Country=="United States of America" | Country
=="Canada",Year %in% c(2008,2009,2010,2011,2012,2013,2014,2015,2016)
) %>%
  ggplot(aes(x=Year,y=Mortality_Rate_Per_1000_Population,color=Gende
r))+
  geom_point()+
  geom_smooth(method = "auto", aes(group=1))+
  facet_wrap(~ Country)+
  scale_y_continuous(breaks=seq(0,300,10))+
  xlab("Year")+
  ylab("Adult mortality rate per 1000 population")+
  labs(size="Health expenditure per capita (PPP)")
```

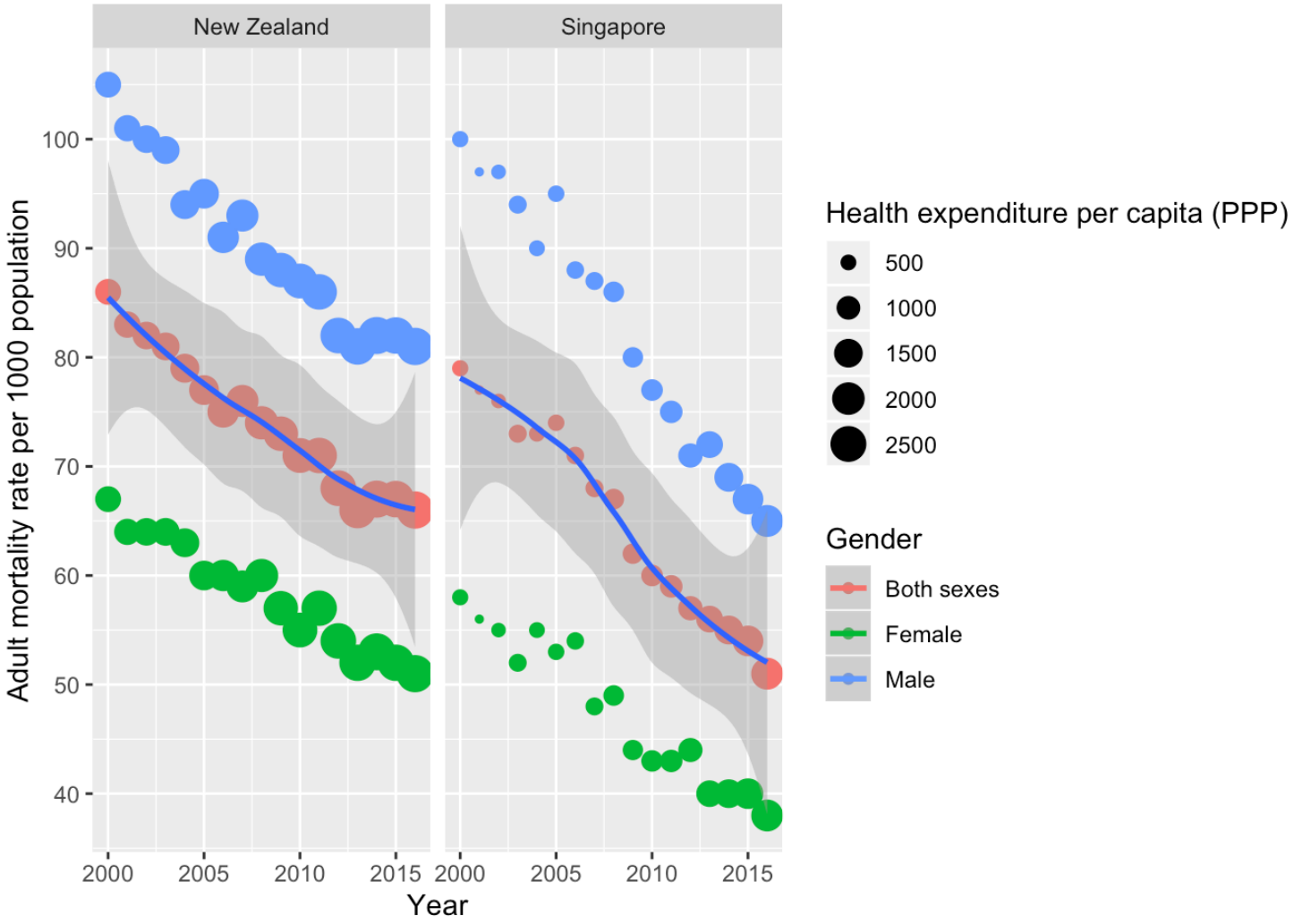
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Answer: For both charts, the smooth lines represent the adult mortality rates of both sexes in Canada and US.The mortality rates of females are less than the mortality rates of males for Cnanada and US. There is the difference.The adult mortality rates of Canada decrease from 2008 to 2016, but the adult mortality rates of US increase slightly from 2008 to 2016, which matches result from the problem 4 plot.Reasons for the increasing adult mortality rates in US are: the crude birth rates show that the live birth rates per 1000 women decrease from 2010, but the drug overdose death rates consistantly increase from 2010, therefore, the overall adult mortality rates of US increase from 2010.

```
#Problem 6
newdataset3 %>% filter(Country=="Singapore" | Country=="New Zealand"
) %>%
  ggplot(aes(x=Year,y=Mortality_Rate_Per_1000_Population,color=Gende
r))+
  geom_point(aes(size=Values))+
  geom_smooth(method = "auto", aes(group=1))+
  facet_wrap(~ Country)+
  scale_y_continuous(breaks=seq(0,200,10))+
  xlab("Year")+
  ylab("Adult mortality rate per 1000 population")+
  labs(size="Health expenditure per capita (PPP)")
```

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



Answer: For both charts, the smooth lines represent the adult mortality rates of both sexes in New Zealand and Singapore.The mortality rates of females are less than the mortality rates of males for both countries. Expenditure on health measures the final

consumption of health goods and services.By adding the size aesthetic to visualize the health expenditure, we can see that both New Zealand and Singapore increase the expenditure on health from 2000 to 2016, and the adult mortality rates for both countries decrease based on that.