

# LDA final project

*Kaitlin Maciejewski, Morgan de Ferrante, Bingnan Li, Volha Tryputsen*

## Exploratory

```
data <- read.csv("../final_data/frmgham2.csv") %>% clean_names()  
attach(data)
```

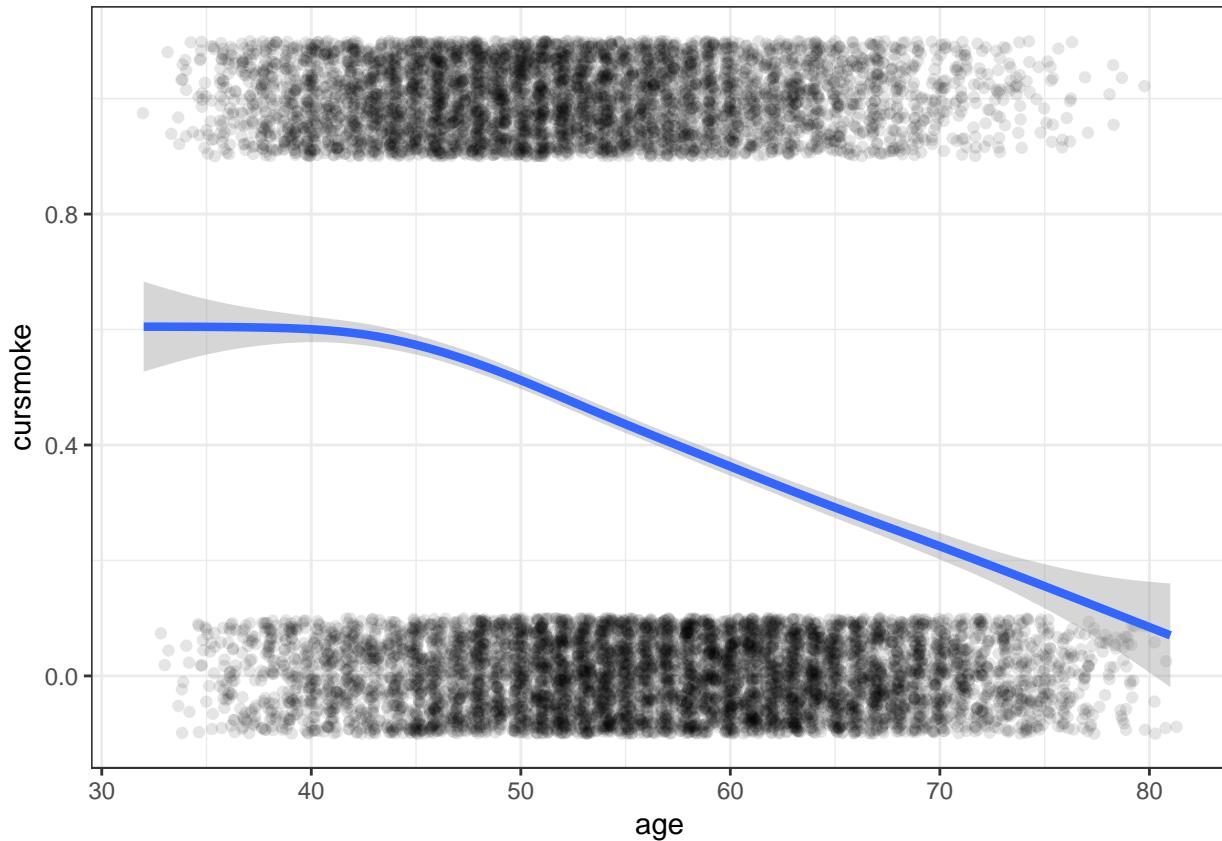
### Smoking vs. Age, Sex

(1): Smoking ~ age, sex

Is there a relationship between age and smoking status?

ANS: Yes, the proportion of smokers decreases with the age.

```
data %>%  
  select(cursmoke, age) %>%  
  ggplot(aes(x = age, y = cursmoke)) +  
  geom_jitter(height = 0.1, alpha = 0.1) +  
  geom_smooth(lwd = 1.5) +  
  theme_bw()  
  
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

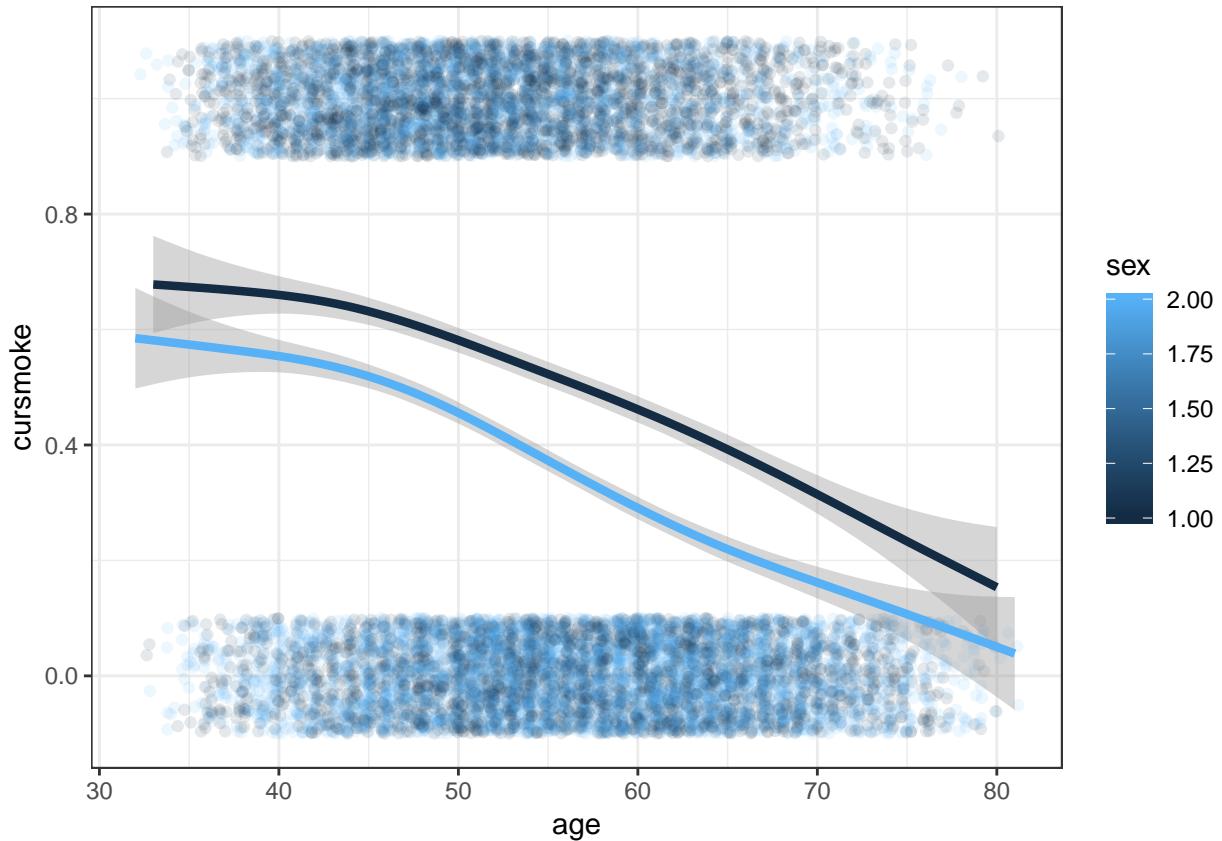


**Does this relationship differ by sex?**

ANS: There is a higher proportion of smoker among men compared to women as both age ,but there is no interaction between age and sex.

```
data %>%
  select(cursmoke, age, sex) %>%
  ggplot(aes(x = age, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

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



## (2) number of cigarettes ~ age, sex

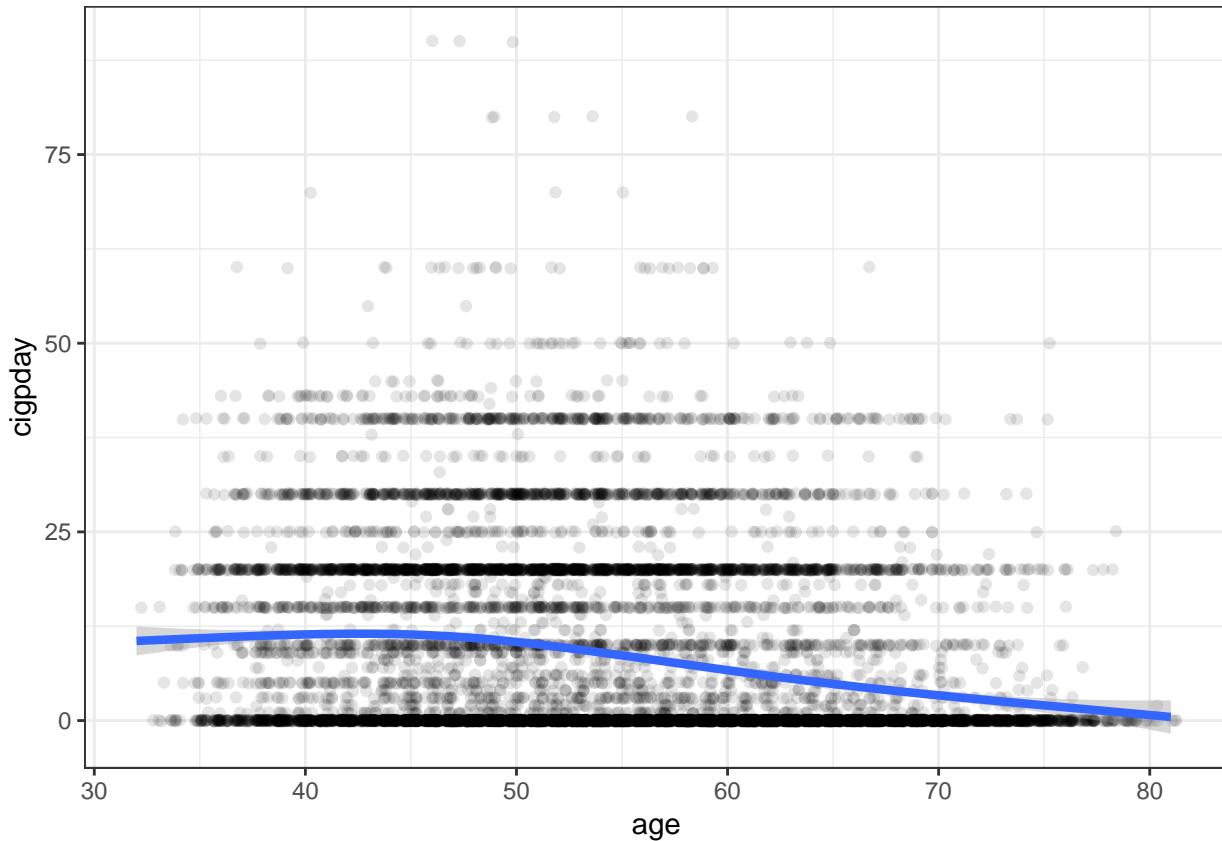
Is there a relationship between the number of cigarettes smoked per day and age?

ANS: Yes, number of sigarets smoked per day stays constant for 30-50 years old and decreases with age after 50 years old.

All

```
data %>%
  select(cigpday, age) %>%
  ggplot(aes(x = age, y = cigpday)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

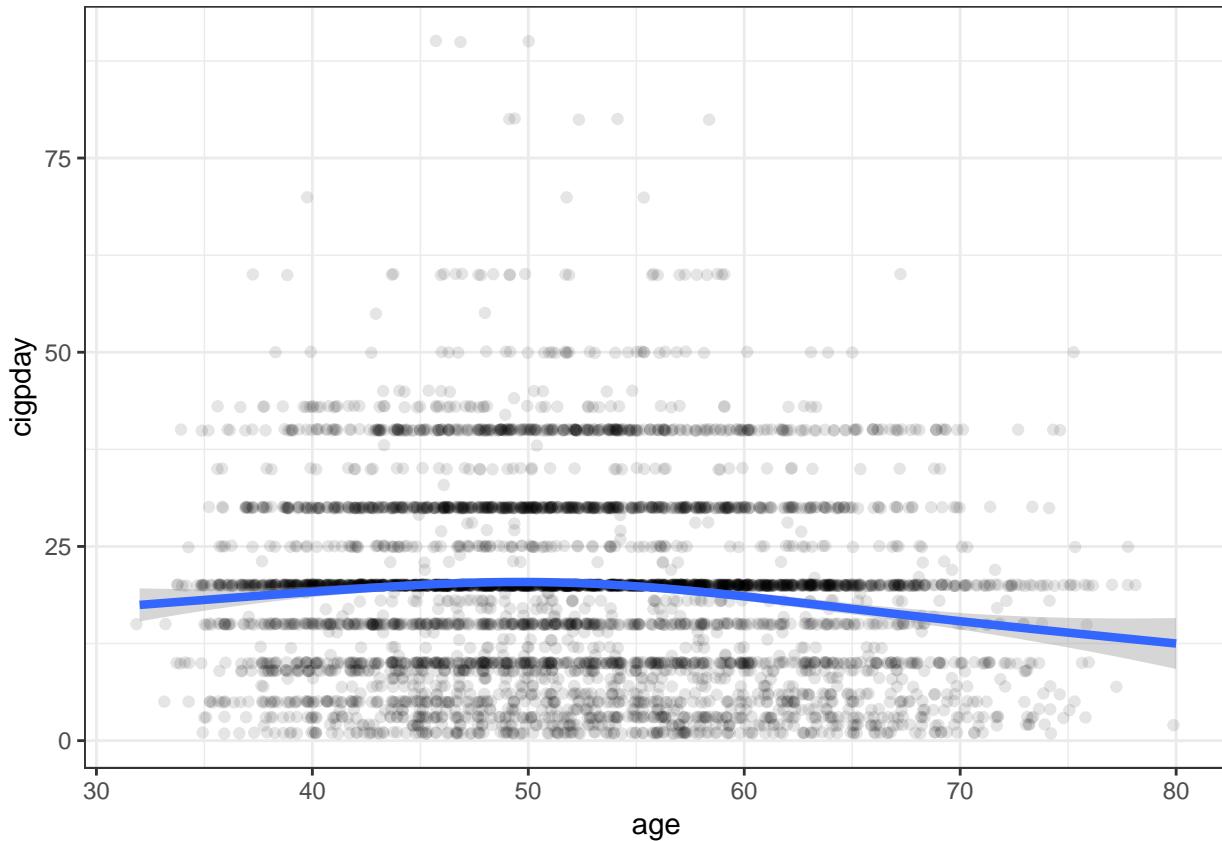
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## Warning: Removed 79 rows containing non-finite values (stat_smooth).
## Warning: Removed 79 rows containing missing values (geom_point).
```



### Smokers only

```
data %>%
  select(cigpday, age) %>%
  filter(cigpday > 0) %>%
  ggplot(aes(x = age, y = cigpday)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

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



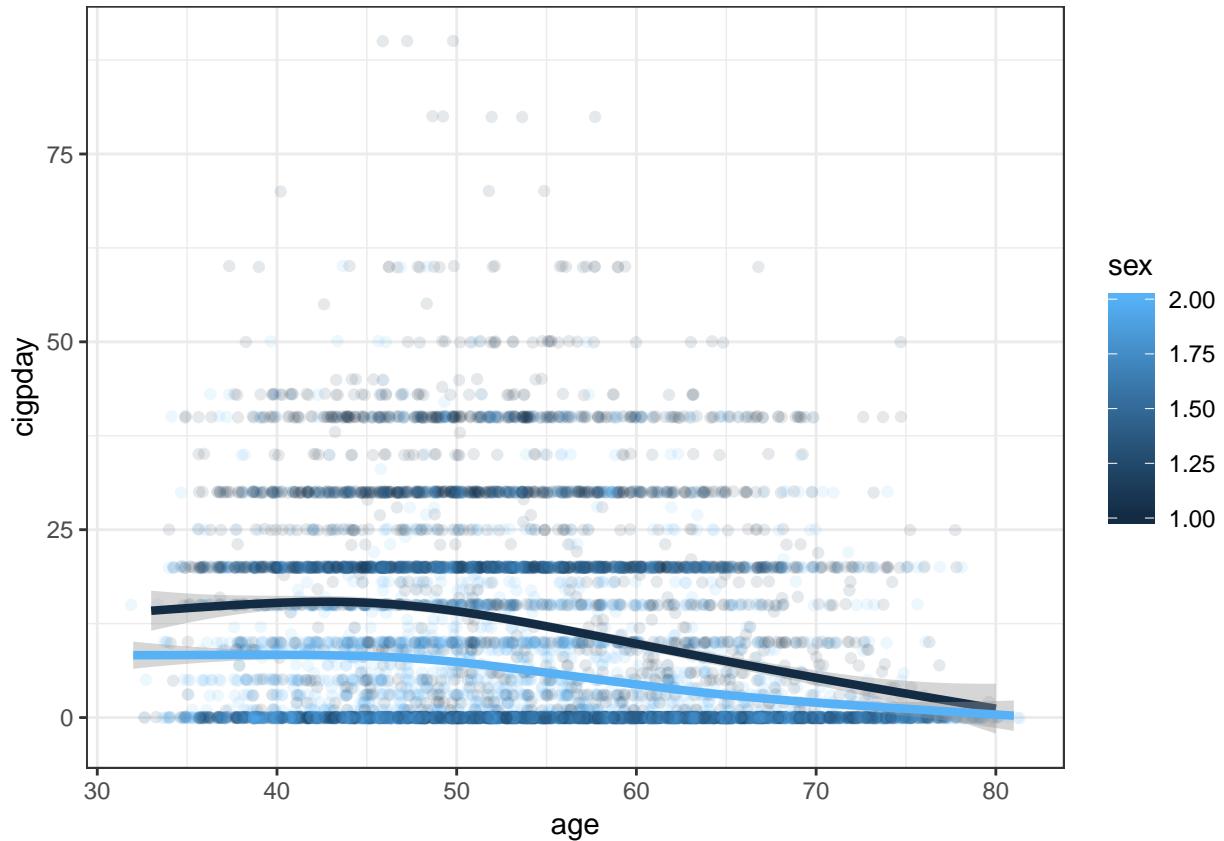
**Does this relationship differ by sex?**

ANS: There is sex effect (men smoke higher number of sigarets per day than women across age), but there is no sex and age interaction.

All

```
data %>%
  select(cigpday, age, sex) %>%
  ggplot(aes(x = age, y = cigpday, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

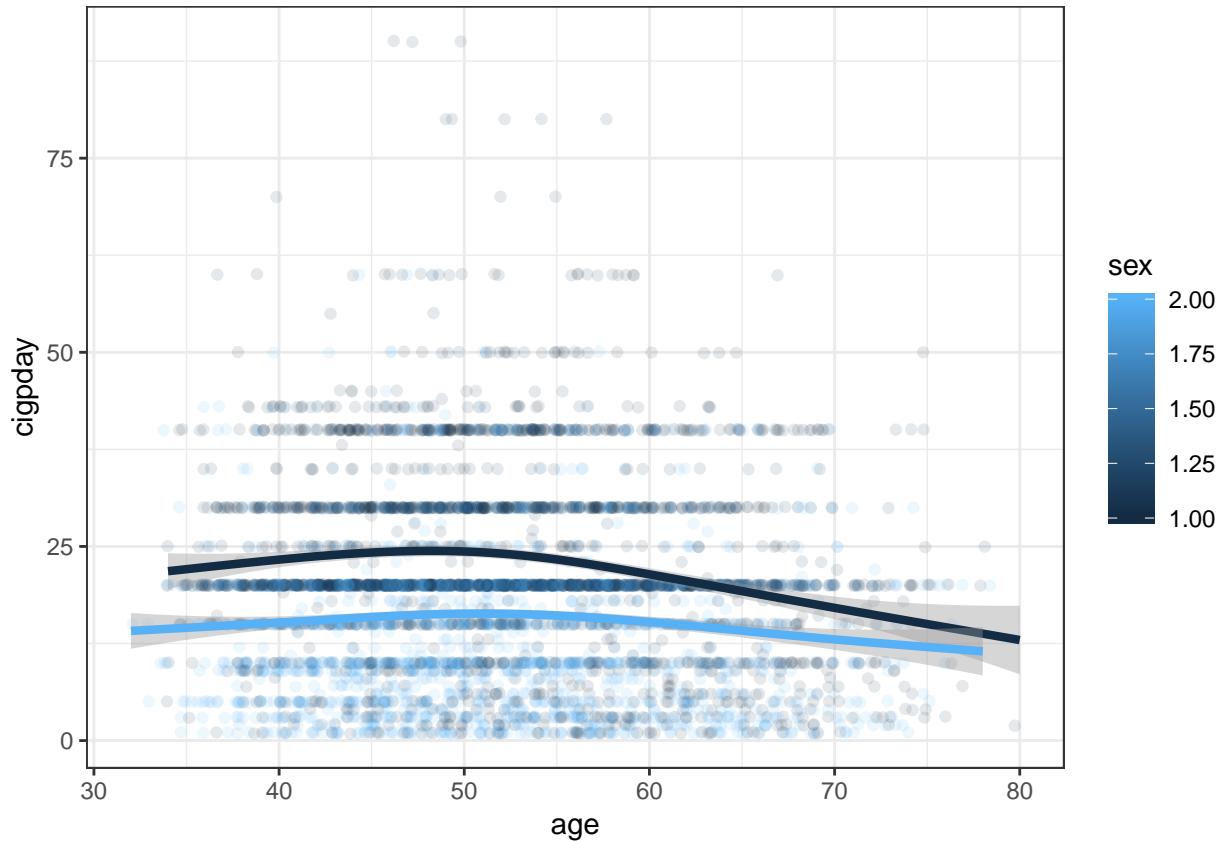
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## Warning: Removed 79 rows containing non-finite values (stat_smooth).
## Warning: Removed 79 rows containing missing values (geom_point).
```



### Smokers only

```
data %>%
  select(cigpday, age, sex) %>%
  filter(cigpday > 0) %>%
  ggplot(aes(x = age, y = cigpday, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

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



## Smoking vs. health outcomes

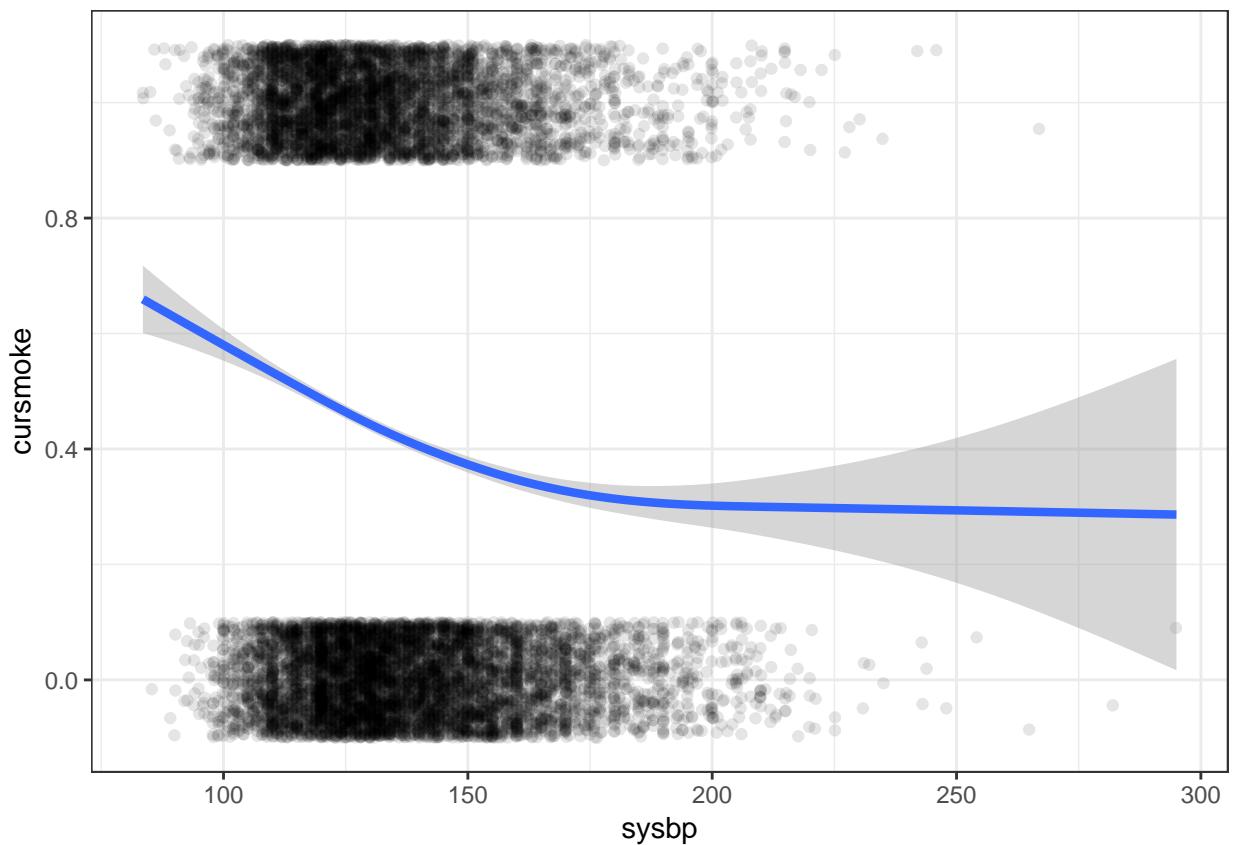
(1) The relationship between current smoking status and systolic blood pressure.

`smoking ~ sysbp`

ANS: Proportion of smokers decreases with increase of systolic blood pressure

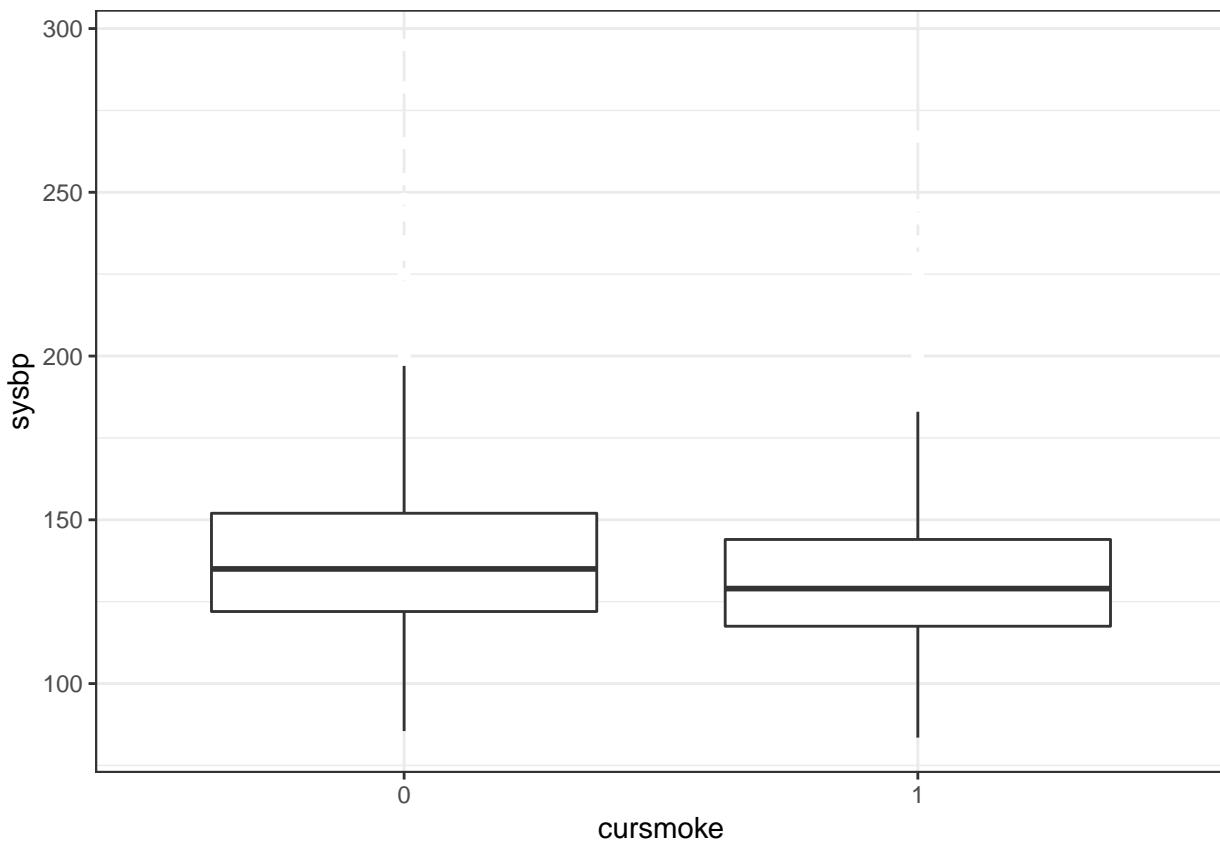
```
data %>%
  select(cursmoke, sysbp) %>%
  ggplot(aes(x = sysbp, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

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



ANS: slightly higher sysbp for non-smokers

```
data %>%
  select(cursmoke, sysbp) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = sysbp, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

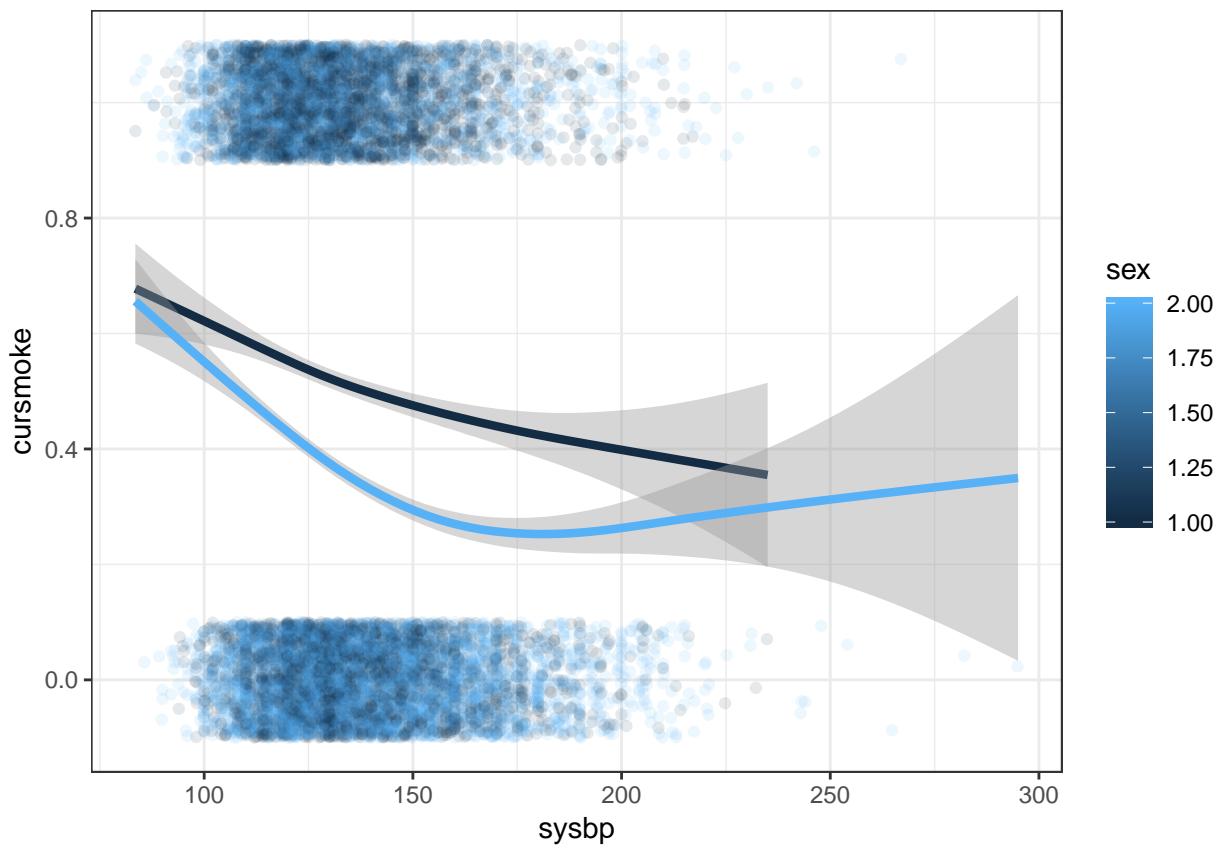


**smoking ~ sysbp, sex**

ANS: Proportion of smokers decreases with increase of systolic blood pressure; the proportion is higher for men (sex effect).

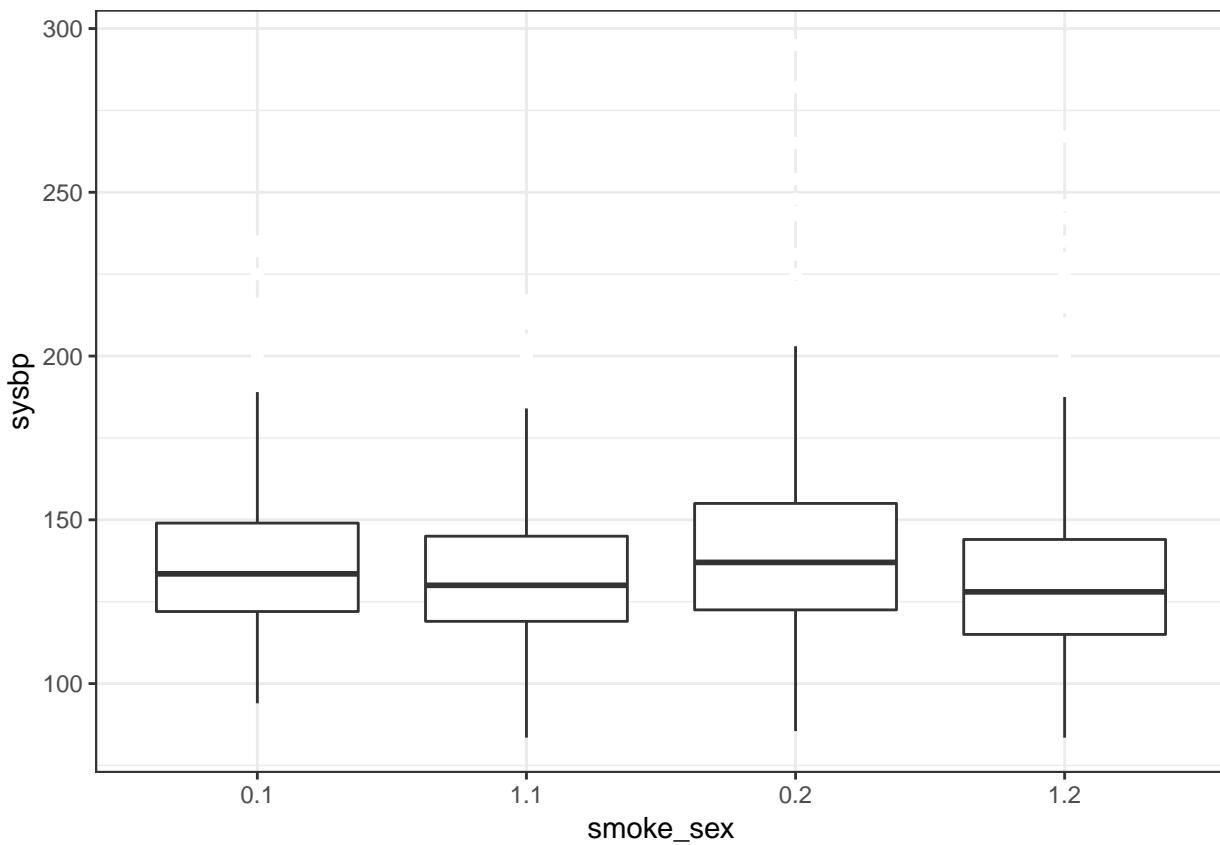
```
data %>%
  select(cursmoke, sysbp, sex) %>%
  ggplot(aes(x = sysbp, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

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



ANS: no differences in sysbp between male and female smokers and non-smokers

```
data %>%
  select(cursmoke, sex, sysbp) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = sysbp, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```



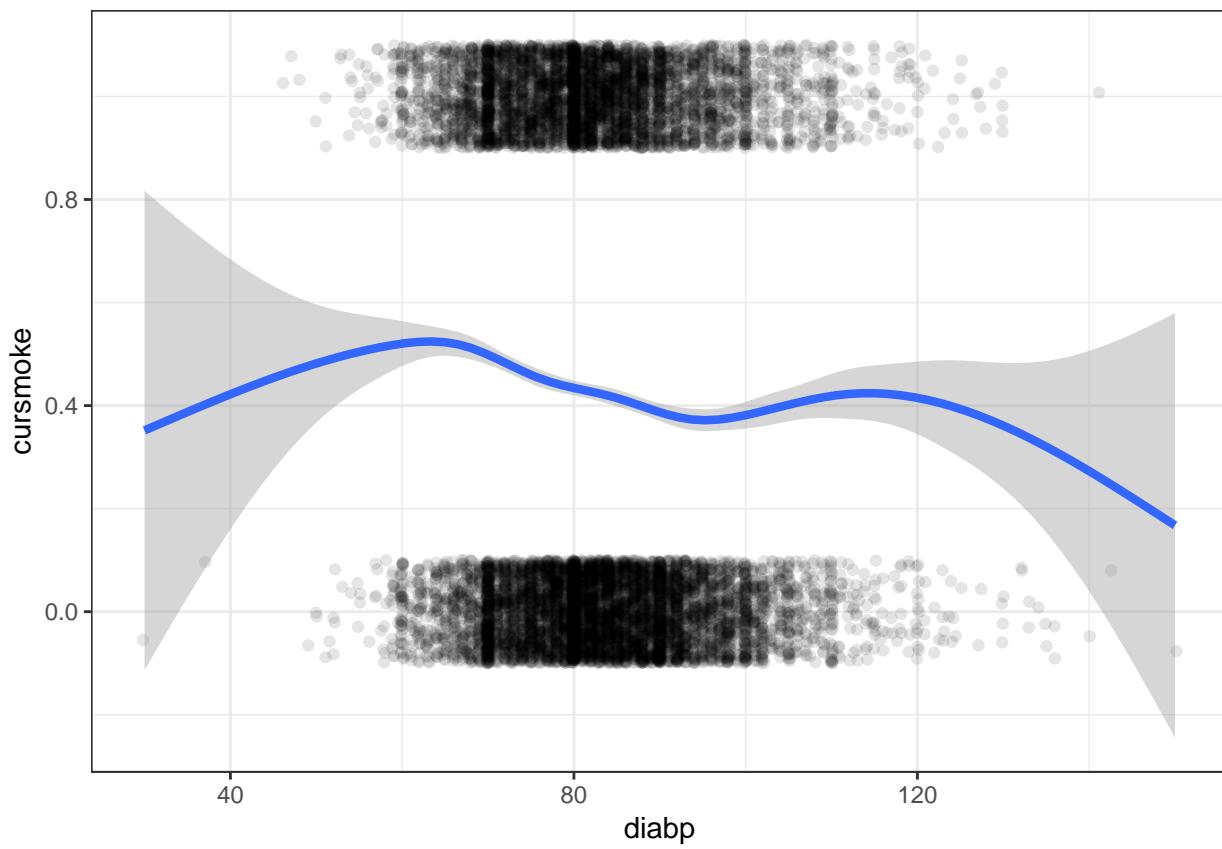
(2) The relationship between current smoking status and diastolic blood pressure.

**smoking ~ diabp**

ANS: Proportion of smokers decreases with increase of diastolic blood pressure for BP=100 ad then proportion increases again (latter could be due to not enough data)

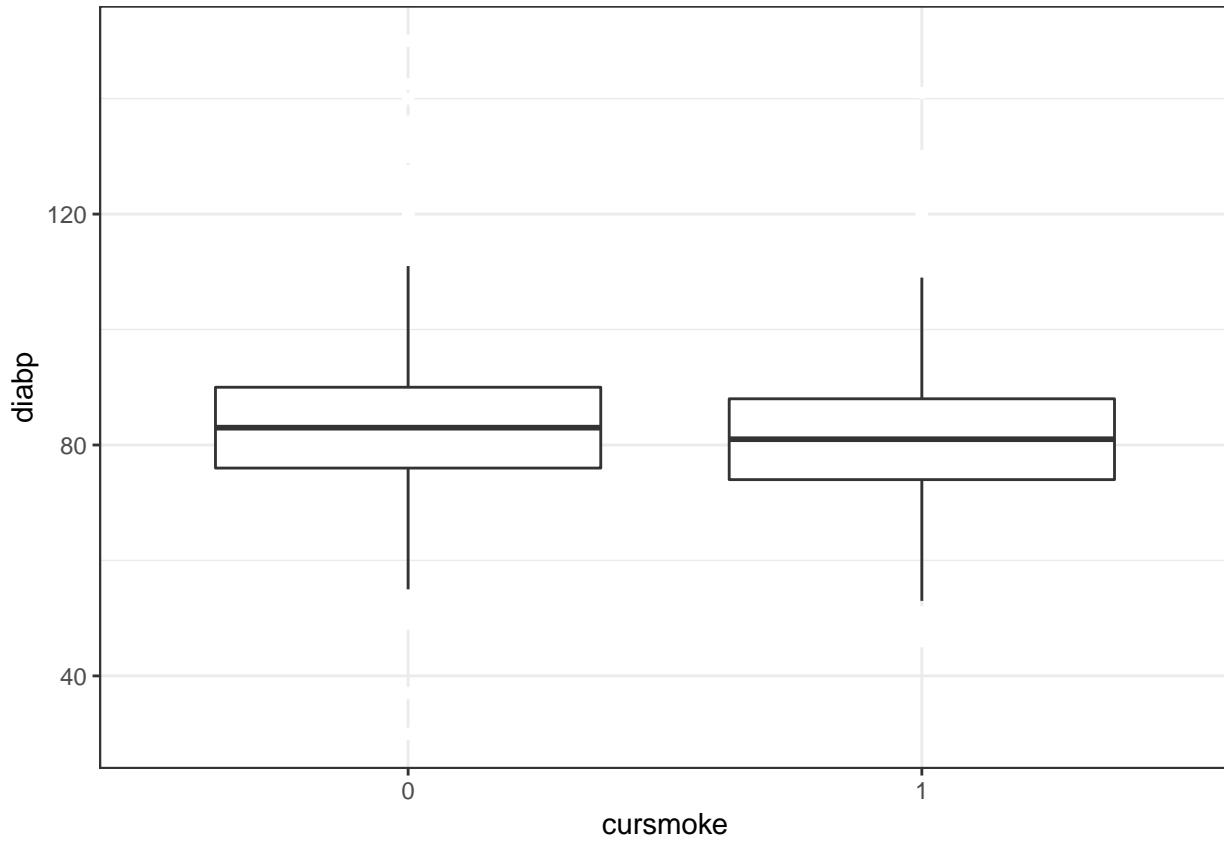
```
data %>%
  select(cursmoke, diabp) %>%
  ggplot(aes(x = diabp, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

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



ANS: no difference

```
data %>%
  select(cursmoke, diabp) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = diabp, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

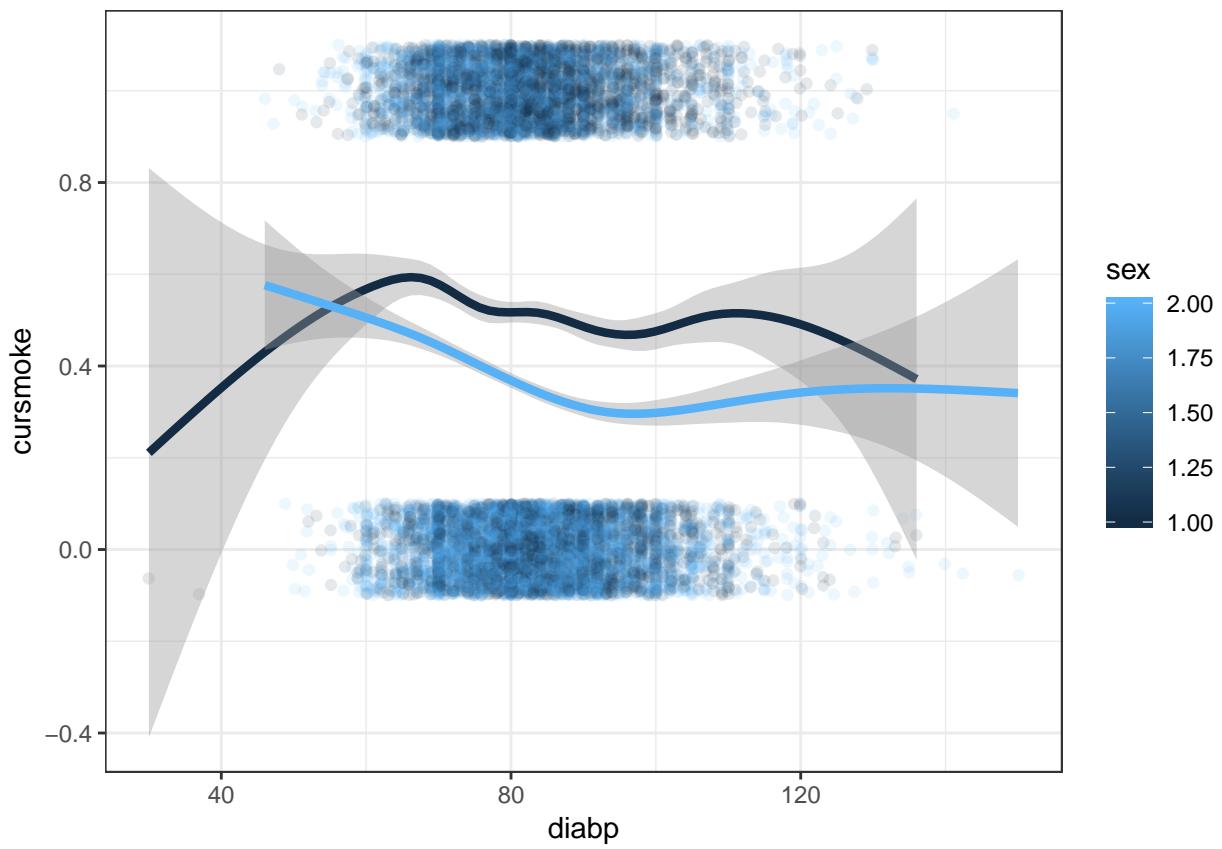


**smoking ~ diabp, sex**

ANS: Proportion of smokers decreases with increase of diastolic blood pressure; the proportions are higher for men (sex effect).

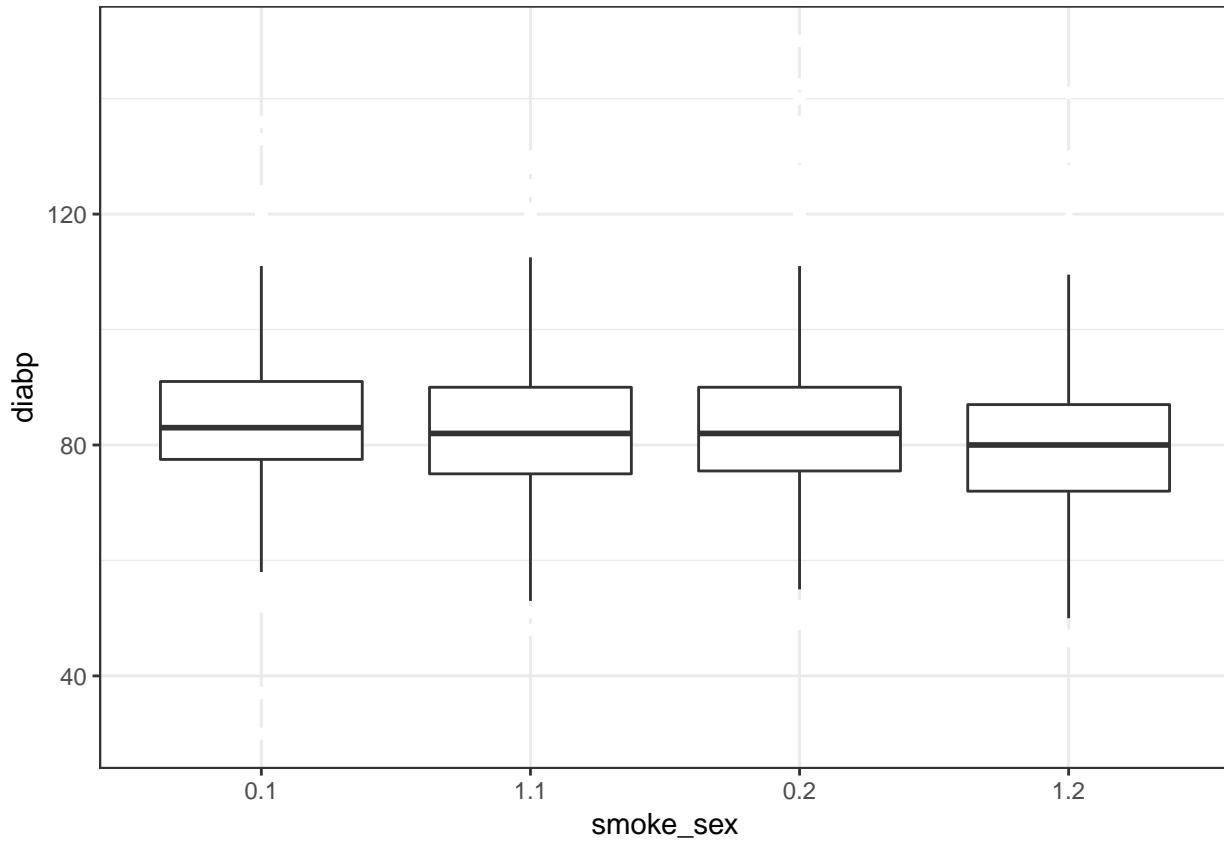
```
data %>%
  select(cursmoke, diabp, sex) %>%
  ggplot(aes(x = diabp, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

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



ANS: no difference

```
data %>%
  select(cursmoke, sex, diabp) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = diabp, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```



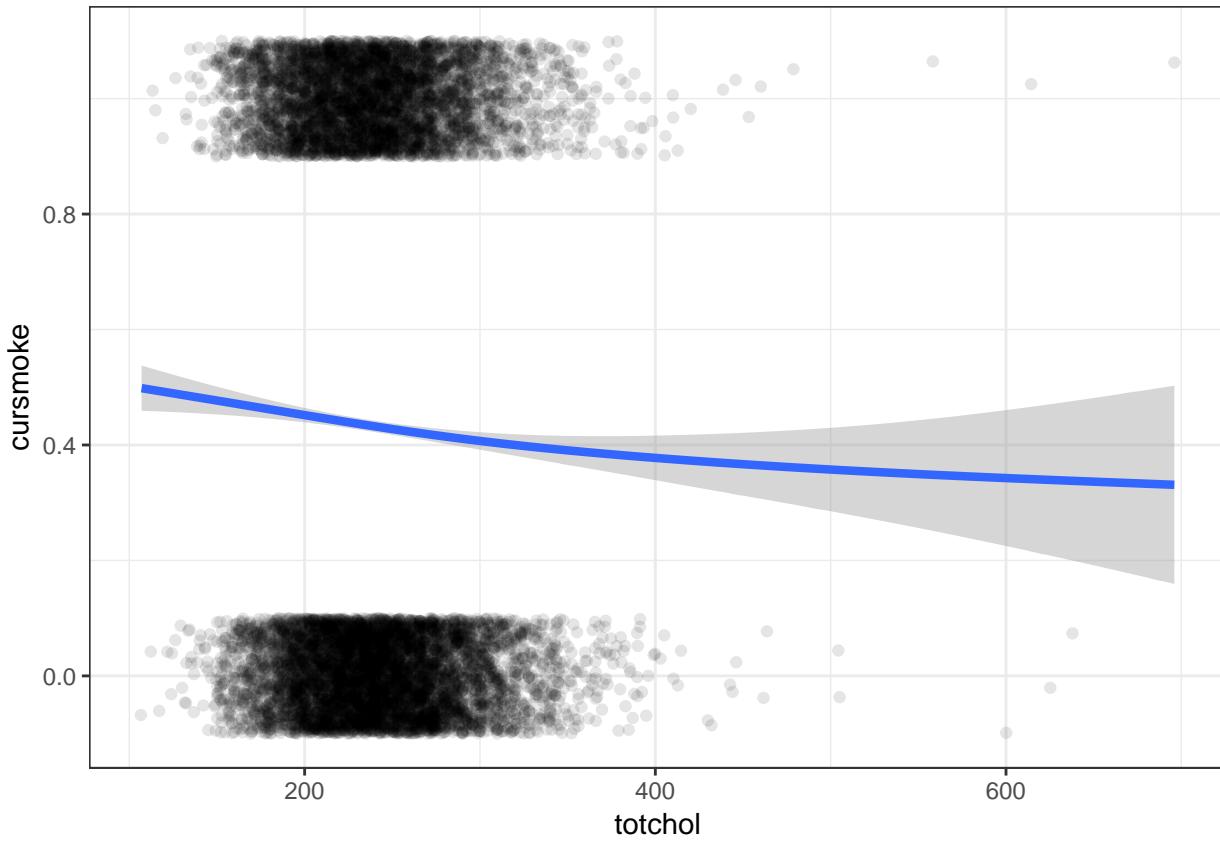
### (3) The relationship between current smoking status and serum total cholesterol.

**smoking ~ totchol**

ANS: Proportion of smokers slightly decreases with increase of total cholesterol values

```
data %>%
  select(cursmoke, totchol) %>%
  ggplot(aes(x = totchol, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

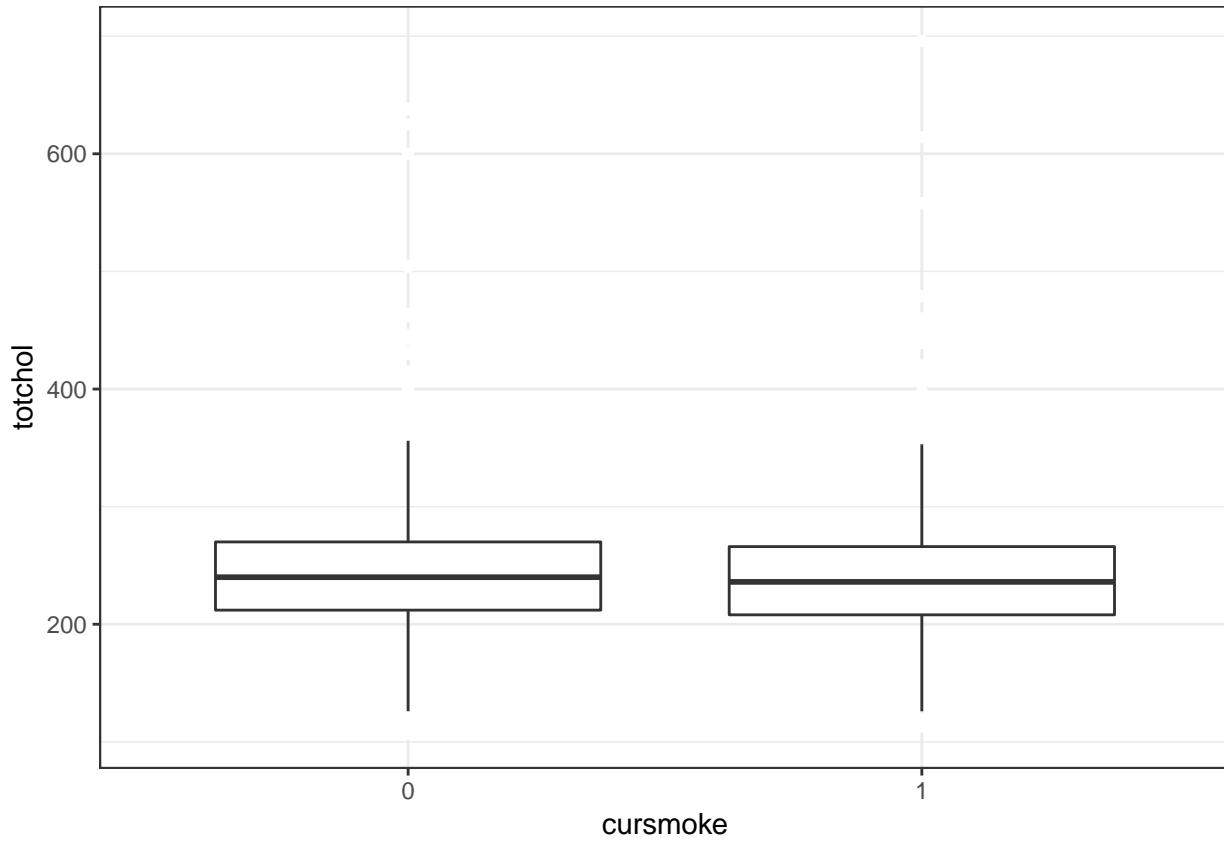
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## Warning: Removed 409 rows containing non-finite values (stat_smooth).
## Warning: Removed 409 rows containing missing values (geom_point).
```



ANS: no difference

```
data %>%
  select(cursmoke, totchol) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = totchol, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()

## Warning: Removed 409 rows containing non-finite values (stat_boxplot).
```

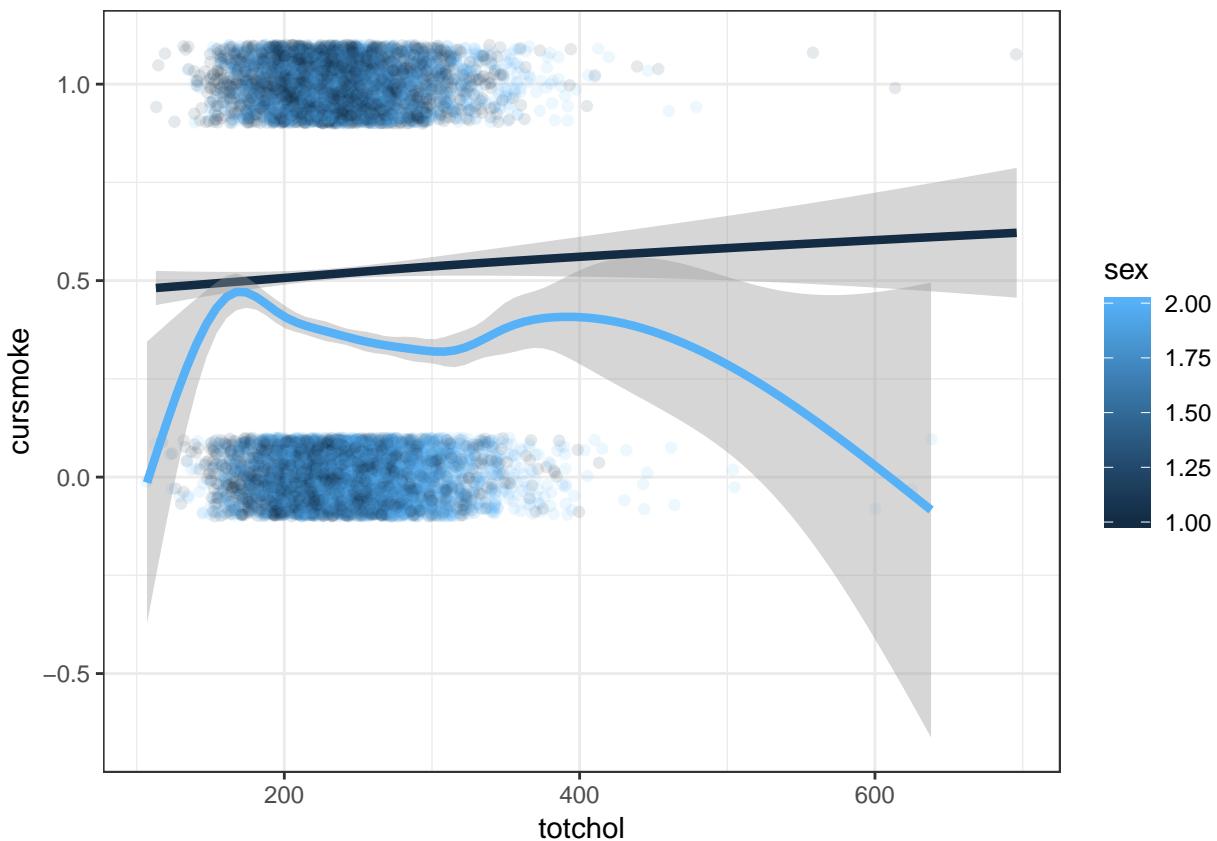


**smoking ~ totchol, sex [!!!]**

ANS: Proportion of smokers has non-linear relationship with total cholesterol for women; proportions increases with increase in total cholesterol for men (sex by totchol interaction effect).

```
data %>%
  select(cursmoke, totchol, sex) %>%
  ggplot(aes(x = totchol, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()

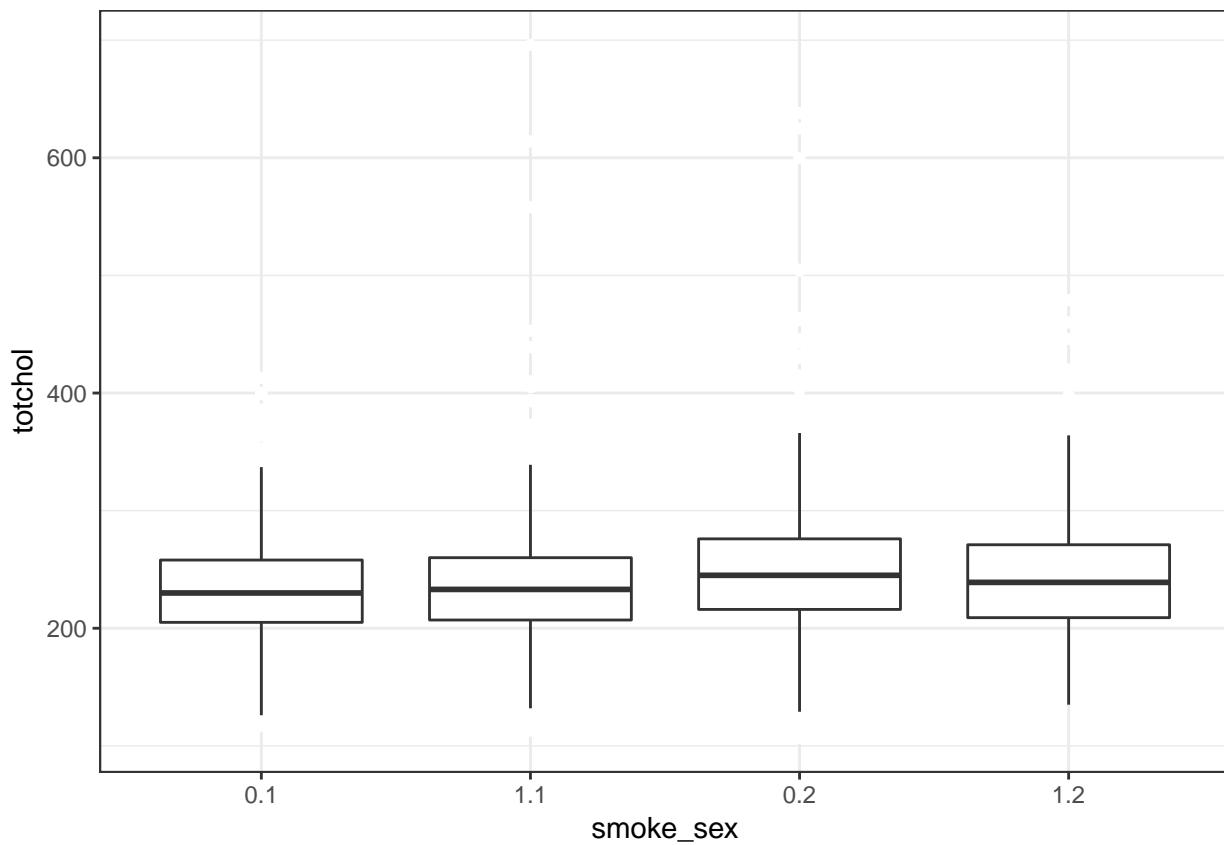
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## Warning: Removed 409 rows containing non-finite values (stat_smooth).
## Warning: Removed 409 rows containing missing values (geom_point).
```



ANS: no difference

```
data %>%
  select(cursmoke, sex, totchol) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = totchol, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

```
## Warning: Removed 409 rows containing non-finite values (stat_boxplot).
```



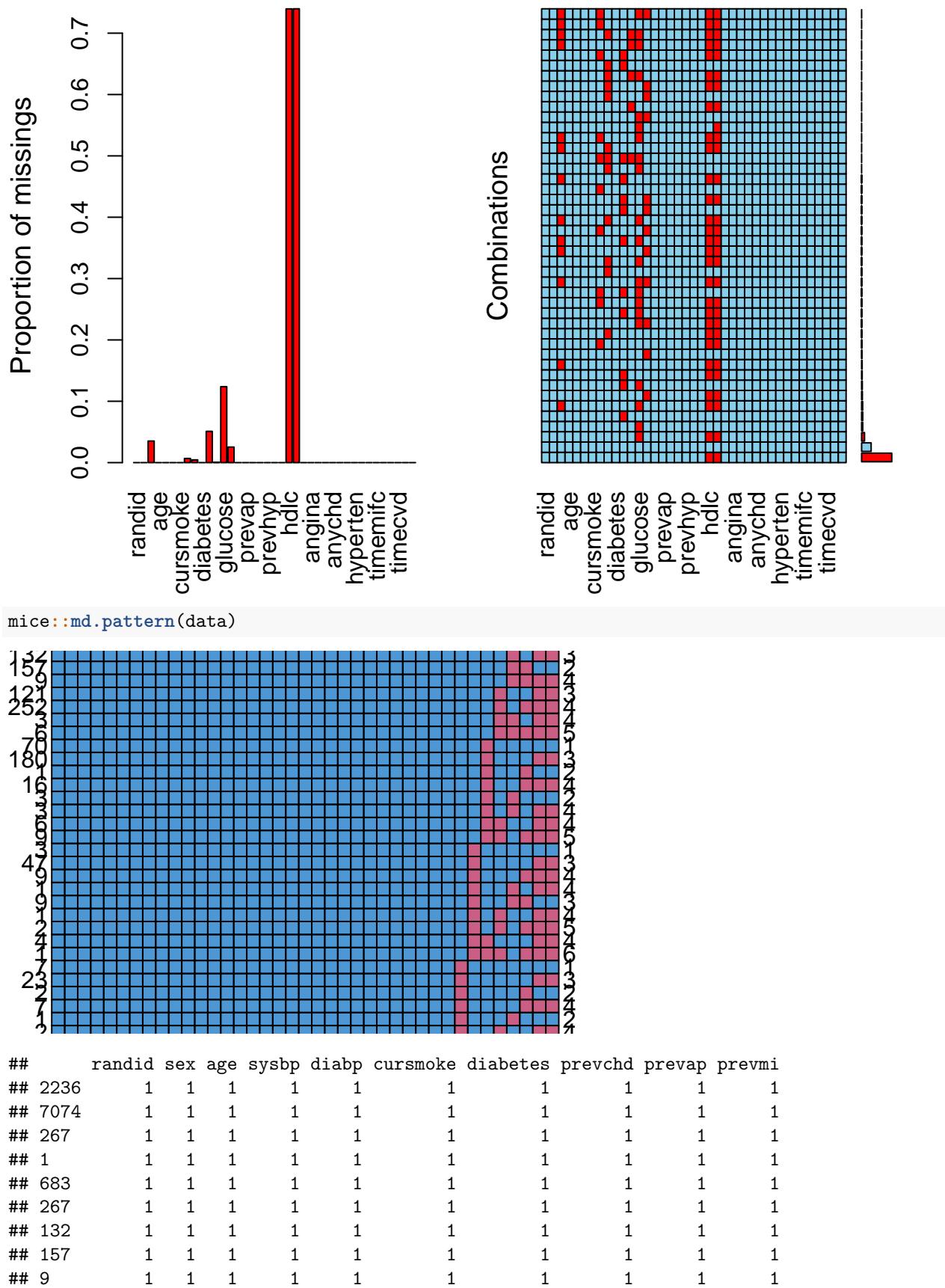
```
library(dplyr)
prop <- round(colSums(is.na(data))/dim(data)[1], 3)

knitr::kable(sort(prop, decreasing = TRUE)[1:9], col.names = "Proportion of NAs")
```

Proportion of NAs	
hdlc	0.740
ldlc	0.740
glucose	0.124
bpmeds	0.051
totchol	0.035
educ	0.025
cigpday	0.007
bmi	0.004
heartrte	0.001

```
#MISSING DATA ANALYSIS
VIM::aggr(data, prop=T, numbers=T)

## Warning in plot.aggr(res, ...): not enough vertical space to display
## frequencies (too many combinations)
```



```

## 121      1 1 1 1 1 1 1 1 1 1
## 252      1 1 1 1 1 1 1 1 1 1
## 3       1 1 1 1 1 1 1 1 1 1
## 6       1 1 1 1 1 1 1 1 1 1
## 70      1 1 1 1 1 1 1 1 1 1
## 180     1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 16      1 1 1 1 1 1 1 1 1 1
## 3       1 1 1 1 1 1 1 1 1 1
## 3       1 1 1 1 1 1 1 1 1 1
## 6       1 1 1 1 1 1 1 1 1 1
## 9       1 1 1 1 1 1 1 1 1 1
## 3       1 1 1 1 1 1 1 1 1 1
## 47      1 1 1 1 1 1 1 1 1 1
## 9       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 9       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 2       1 1 1 1 1 1 1 1 1 1
## 4       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 7       1 1 1 1 1 1 1 1 1 1
## 23      1 1 1 1 1 1 1 1 1 1
## 2       1 1 1 1 1 1 1 1 1 1
## 7       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 2       1 1 1 1 1 1 1 1 1 1
## 4       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 2       1 1 1 1 1 1 1 1 1 1
## 1       0 0 0 0 0 0 0 0 0 0
##          prevstrk prevhyp time period death angina hospmi mi_fchd anychd
## 2236     1 1 1 1 1 1 1 1 1 1
## 7074     1 1 1 1 1 1 1 1 1 1
## 267      1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 683      1 1 1 1 1 1 1 1 1 1
## 267      1 1 1 1 1 1 1 1 1 1
## 132      1 1 1 1 1 1 1 1 1 1
## 157      1 1 1 1 1 1 1 1 1 1
## 9        1 1 1 1 1 1 1 1 1 1
## 121     1 1 1 1 1 1 1 1 1 1
## 252     1 1 1 1 1 1 1 1 1 1
## 3        1 1 1 1 1 1 1 1 1 1
## 6        1 1 1 1 1 1 1 1 1 1
## 70      1 1 1 1 1 1 1 1 1 1
## 180     1 1 1 1 1 1 1 1 1 1
## 1       1 1 1 1 1 1 1 1 1 1
## 16      1 1 1 1 1 1 1 1 1 1

```

```

## 3      1      1      1      1      1      1      1      1      1      1
## 3      1      1      1      1      1      1      1      1      1      1
## 6      1      1      1      1      1      1      1      1      1      1
## 9      1      1      1      1      1      1      1      1      1      1
## 3      1      1      1      1      1      1      1      1      1      1
## 47     1      1      1      1      1      1      1      1      1      1
## 9      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 9      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 2      1      1      1      1      1      1      1      1      1      1
## 4      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 7      1      1      1      1      1      1      1      1      1      1
## 23     1      1      1      1      1      1      1      1      1      1
## 2      1      1      1      1      1      1      1      1      1      1
## 7      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 2      1      1      1      1      1      1      1      1      1      1
## 4      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 1      1      1      1      1      1      1      1      1      1      1
## 2      1      1      1      1      1      1      1      1      1      1
## 0      0      0      0      0      0      0      0      0      0      0
##      stroke cvd hyperten timeap timemi timemifc timechd timestrk timecvd
## 2236    1      1      1      1      1      1      1      1      1      1
## 7074    1      1      1      1      1      1      1      1      1      1
## 267     1      1      1      1      1      1      1      1      1      1
## 1       1      1      1      1      1      1      1      1      1      1
## 683     1      1      1      1      1      1      1      1      1      1
## 267     1      1      1      1      1      1      1      1      1      1
## 132     1      1      1      1      1      1      1      1      1      1
## 157     1      1      1      1      1      1      1      1      1      1
## 9       1      1      1      1      1      1      1      1      1      1
## 121     1      1      1      1      1      1      1      1      1      1
## 252     1      1      1      1      1      1      1      1      1      1
## 3       1      1      1      1      1      1      1      1      1      1
## 6       1      1      1      1      1      1      1      1      1      1
## 70      1      1      1      1      1      1      1      1      1      1
## 180     1      1      1      1      1      1      1      1      1      1
## 1       1      1      1      1      1      1      1      1      1      1
## 16      1      1      1      1      1      1      1      1      1      1
## 3       1      1      1      1      1      1      1      1      1      1
## 3       1      1      1      1      1      1      1      1      1      1
## 6       1      1      1      1      1      1      1      1      1      1
## 9       1      1      1      1      1      1      1      1      1      1
## 3       1      1      1      1      1      1      1      1      1      1
## 47     1      1      1      1      1      1      1      1      1      1
## 9       1      1      1      1      1      1      1      1      1      1
## 1       1      1      1      1      1      1      1      1      1      1

```

```

## 9      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 2      1   1     1   1     1   1     1   1     1   1
## 4      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 7      1   1     1   1     1   1     1   1     1   1
## 23     1   1     1   1     1   1     1   1     1   1
## 2      1   1     1   1     1   1     1   1     1   1
## 7      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 2      1   1     1   1     1   1     1   1     1   1
## 4      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 1      1   1     1   1     1   1     1   1     1   1
## 2      1   1     1   1     1   1     1   1     1   1
## 0      0     0     0     0     0     0     0     0     0
##      timedeth timehyp heartrte bmi cigpdday educ totchol bpmeds glucose hdlc
## 2236    1     1     1   1     1   1     1   1     1   1
## 7074    1     1     1   1     1   1     1   1     1   0
## 267     1     1     1   1     1   1     1   1     0   1
## 1      1     1     1   1     1   1     1   1     0   1
## 683     1     1     1   1     1   1     1   1     0   0
## 267     1     1     1   1     1   1     1   0     1   1
## 132     1     1     1   1     1   1     1   0     1   0
## 157     1     1     1   1     1   1     1   0     0   1
## 9       1     1     1   1     1   1     1   0     0   0
## 121     1     1     1   1     1   1     0   1     1   0
## 252     1     1     1   1     1   1     0   1     0   0
## 3       1     1     1   1     1   1     0   0     1   0
## 6       1     1     1   1     1   1     0   0     0   0
## 70      1     1     1   1     1   0     1   1     1   1
## 180     1     1     1   1     1   0     1   1     1   0
## 1       1     1     1   1     1   0     1   1     0   1
## 16      1     1     1   1     1   0     1   1     0   0
## 3       1     1     1   1     1   0     1   0     1   1
## 3       1     1     1   1     1   0     1   0     1   0
## 6       1     1     1   1     1   0     0     1   1   0
## 9       1     1     1   1     1   0     0     1   0   0
## 3       1     1     1   1     0   1     1   1     1   1
## 47      1     1     1   1     0   1     1   1     1   0
## 9       1     1     1   1     0   1     1   1     0   0
## 1       1     1     1   1     0   1     1   0     1   0
## 9       1     1     1   1     0   1     1   0     0   1
## 1       1     1     1   1     0   1     0     1   1   0
## 2       1     1     1   1     0   1     0     1   0   0
## 4       1     1     1   1     0   0     1     1   1   0
## 1       1     1     1   1     0   0     0     0     1   0
## 7       1     1     1   0     1   1     1   1     1   1
## 23      1     1     1   0     1   1     1   1     1   0
## 2       1     1     1   0     1   1     1   1     0   1

```

```

## 7      1      1      1      0      1      1      1      1      0      0
## 1      1      1      1      0      1      1      1      0      1      1
## 2      1      1      1      0      1      1      0      1      1      0
## 4      1      1      1      0      1      1      0      1      0      0
## 1      1      1      1      0      1      0      1      1      1      1
## 1      1      1      1      0      1      0      1      1      1      0
## 1      1      1      0      1      1      1      1      1      1      0
## 1      1      1      0      1      1      0      1      0      0      0
## 1      1      1      0      0      1      1      1      0      0      0
## 1      1      1      0      0      1      1      0      1      0      0
## 2      1      1      0      0      0      1      1      1      0      1
##                      0      0      6      52     79     295     409     593    1440   8600
##      ldlc
## 2236    1      0
## 7074    0      2
## 267     1      1
## 1       0      2
## 683     0      3
## 267     1      1
## 132     0      3
## 157     1      2
## 9       0      4
## 121     0      3
## 252     0      4
## 3       0      4
## 6       0      5
## 70      1      1
## 180     0      3
## 1       1      2
## 16      0      4
## 3       1      2
## 3       0      4
## 6       0      4
## 9       0      5
## 3       1      1
## 47      0      3
## 9       0      4
## 1       0      4
## 9       1      3
## 1       0      4
## 2       0      5
## 4       0      4
## 1       0      6
## 7       1      1
## 23      0      3
## 2       1      2
## 7       0      4
## 1       1      2
## 2       0      4
## 4       0      5
## 1       1      2
## 1       0      4
## 1       0      3
## 1       0      5

```

```

## 1      0      5
## 1      0      6
## 2      1      5
##     8601 20075
# HDLC and LDLC have a lot of missing data, and these variables should be eliminated

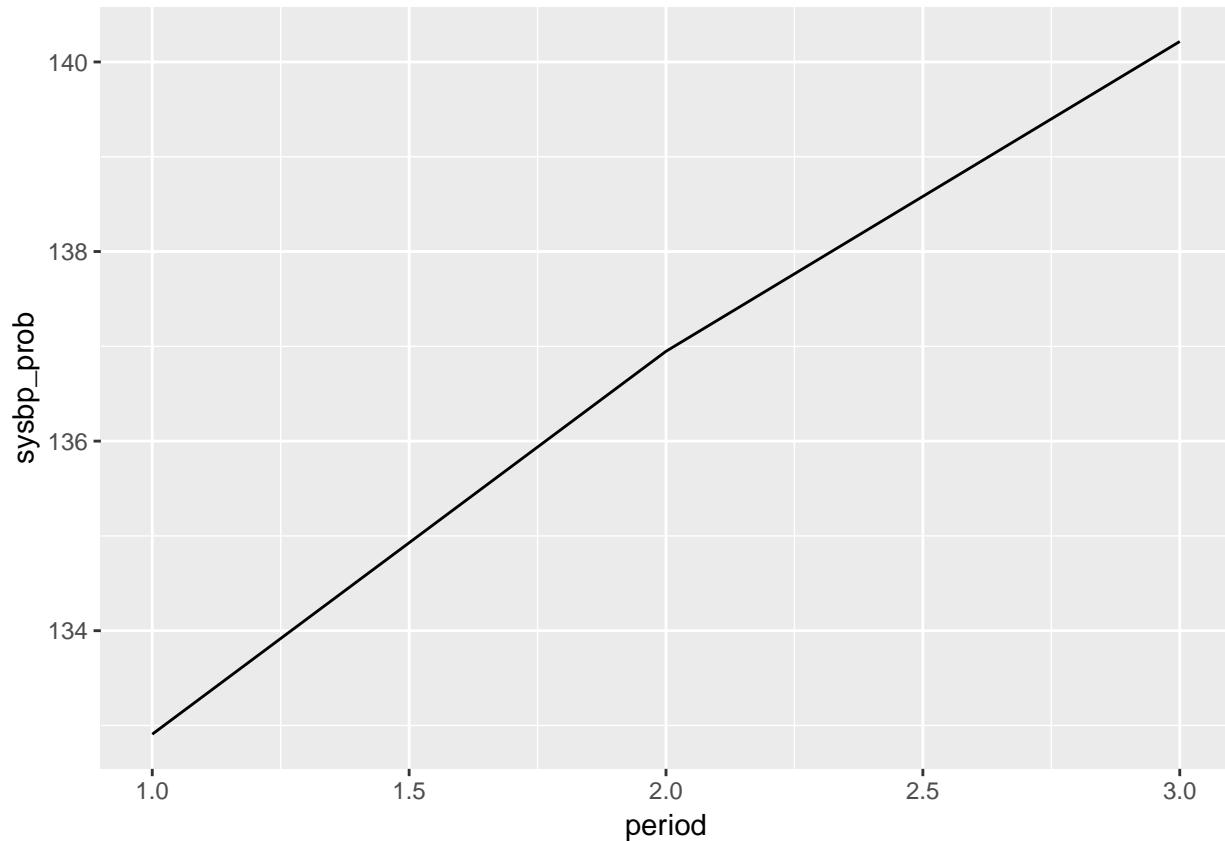
prob.data <- data %>%
  group_by(period) %>%
  summarise(sysbp_prob = sum(sysbp, na.rm = TRUE)/n())

table(data$period)

##
##      1    2    3
## 4434 3930 3263

##plot these probabilities
ggplot(prob.data, aes(y=sysbp_prob,
                      x=period)) +
  geom_line() +
  labs(color = 'Avg Systolic BP')

```



```

prob.data <- data %>%
  group_by(period) %>%
  summarise(diabp_prob = sum(diabp, na.rm = TRUE)/n())

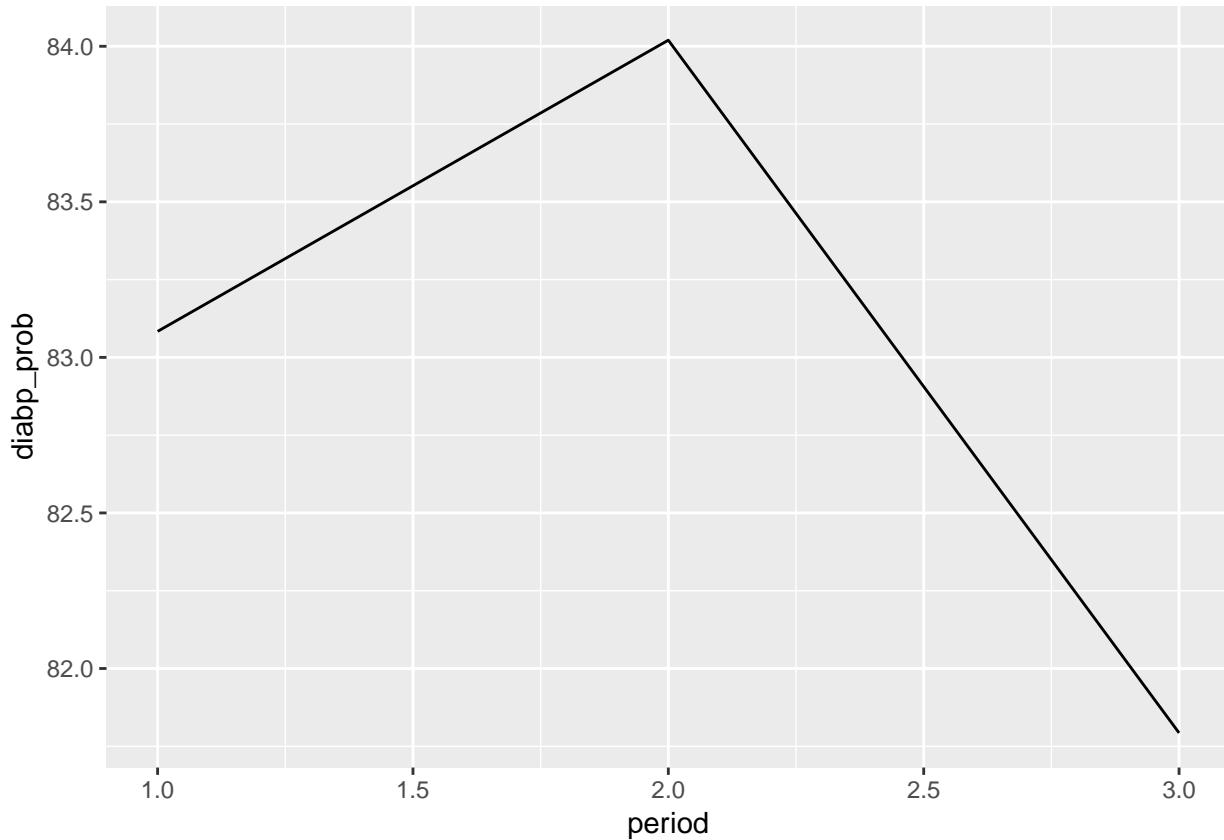
##plot these probabilities
ggplot(prob.data, aes(y=diabp_prob,

```

```

        x=period)) +
geom_line() +
labs(color = 'Avg Diastolic BP')

```

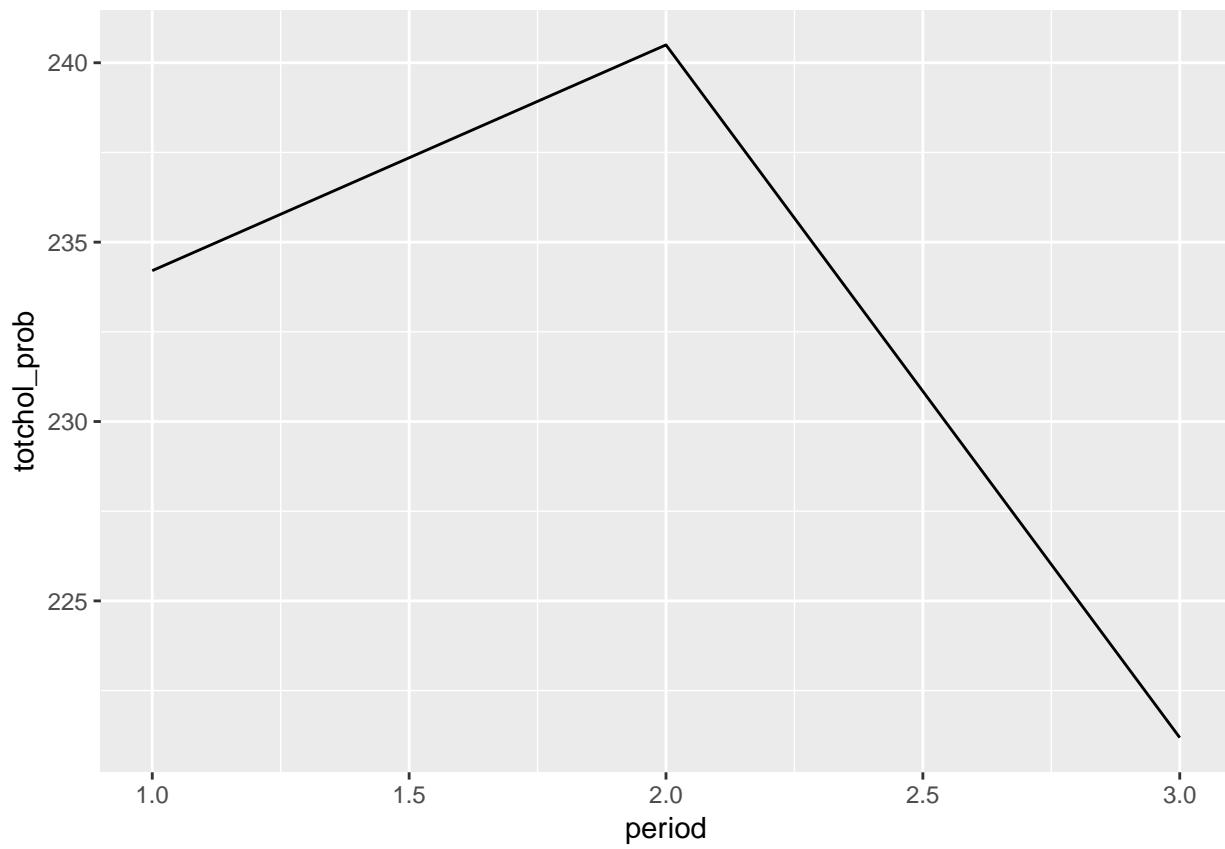


```

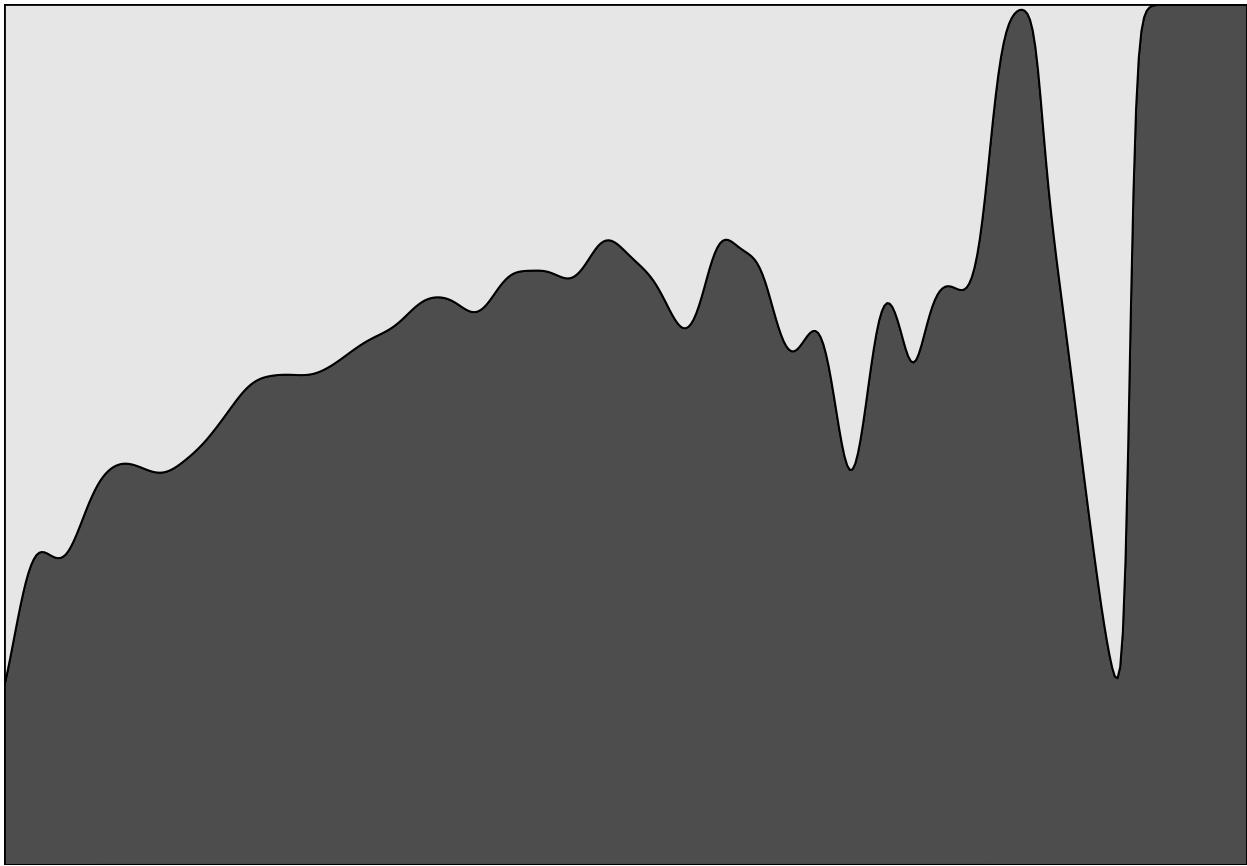
prob.data <- data %>%
  group_by(period) %>%
  summarise(totchol_prob = sum(totchol, na.rm = TRUE)/n())

##plot these probabilities
ggplot(prob.data, aes(y=totchol_prob,
                      x=period)) +
  geom_line() +
  labs(color = 'Avg Total Chol BP')

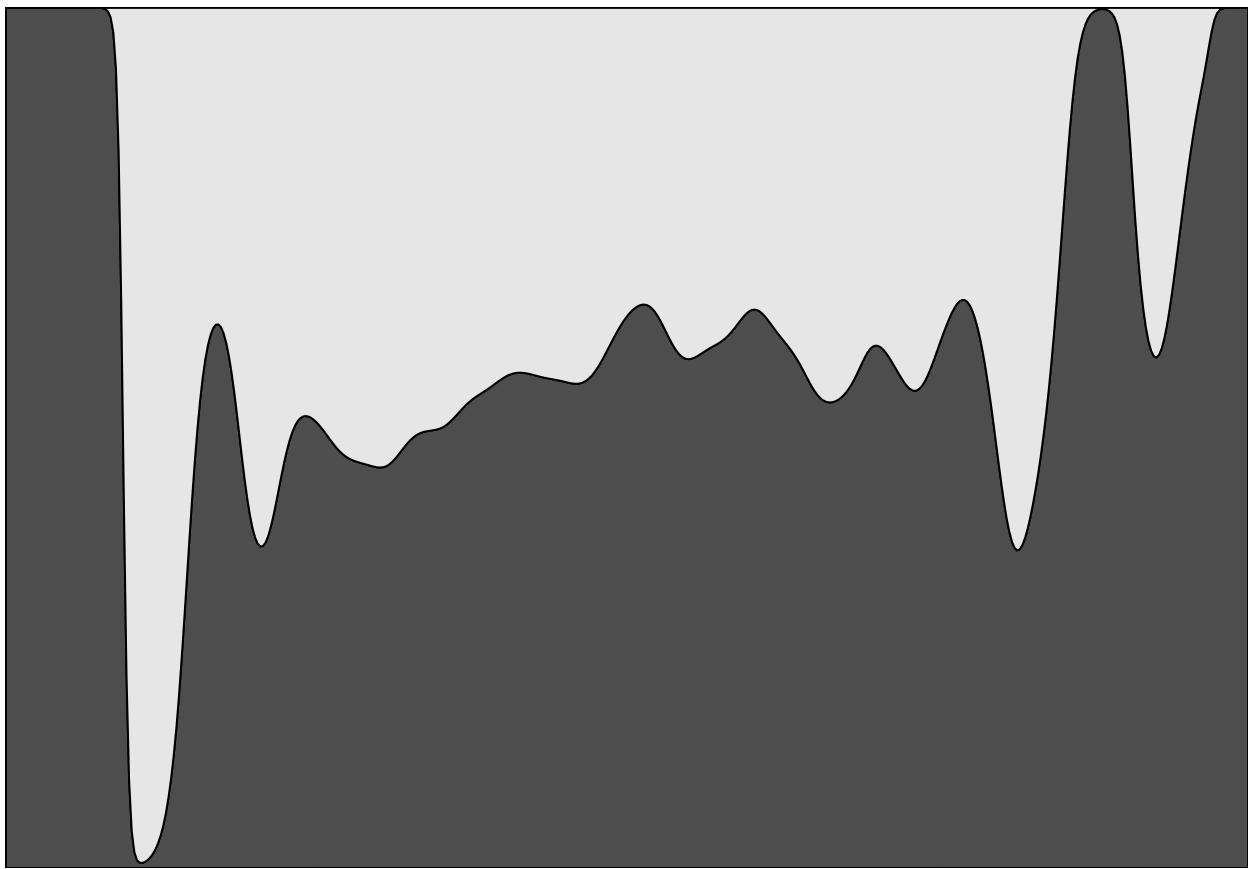
```



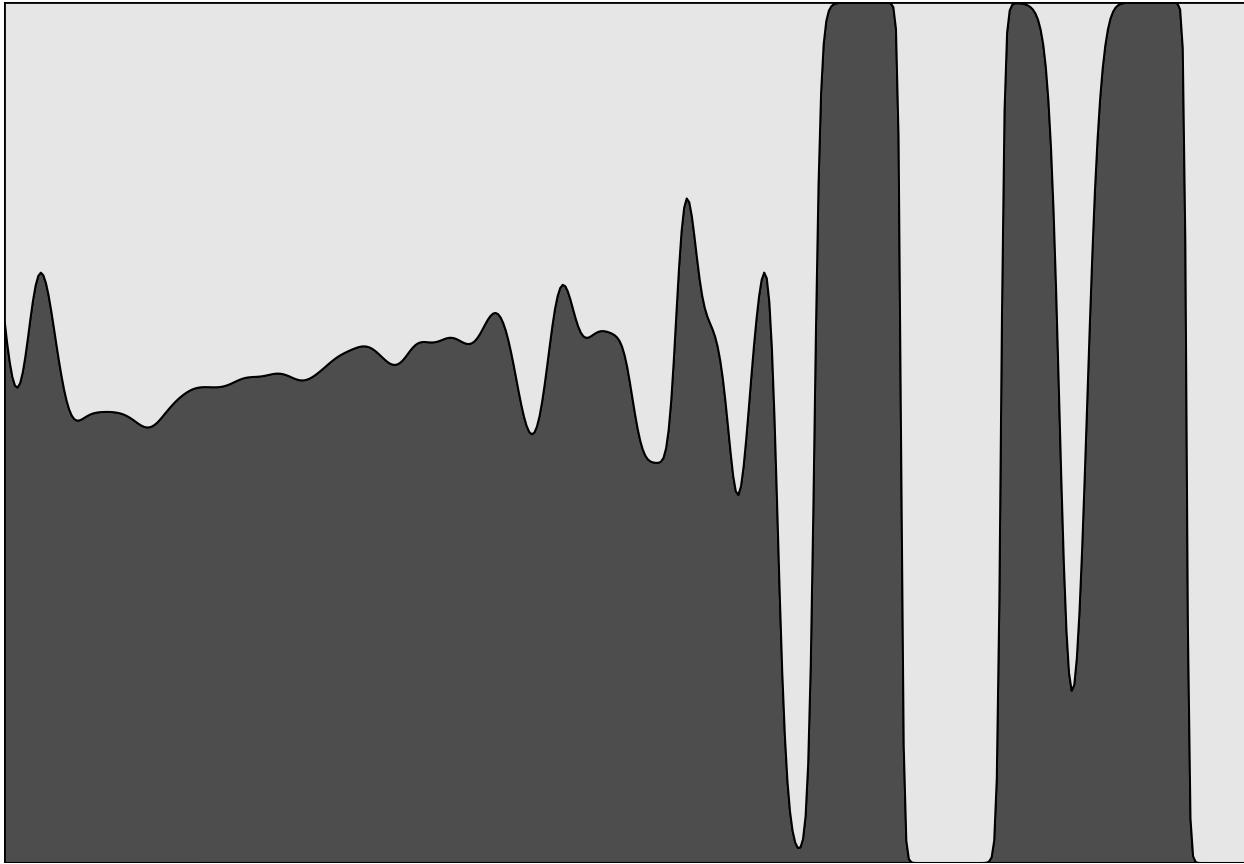
```
cdplot(factor(cursmoke) ~ sysbp, data=data,
      ylab = "Current Smoking Status",
      xlab = "Systolic BP")
```



```
cdplot(factor(cursmoke) ~ diabp, data=data,
      ylab = "Current Smoking Status",
      xlab = "Diatolic BP")
```



```
cdplot(factor(cursmoke) ~ totchol, data=data,
      ylab = "Current Smoking Status",
      xlab = "Total Cholesterol")
```



## models

```
library(lme4)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following object is masked from 'package:tidyverse':
##      expand
library(dplyr)

model_1<-glmer(cursmoke~ age + sex + sysbp + diabp+ totchol + as.factor(educ) + (1|randid), family=binom

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 1.19128 (tol
## = 0.001, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly uniden
## - Rescale variables?;Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?
```

```
knitr::kable(summary(model_1)$coefficients,digits = 3)
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	16.730	0.965	17.332	0.000
age	-0.206	0.010	-20.134	0.000
sex	-3.760	0.324	-11.597	0.000
sysbp	-0.001	0.005	-0.193	0.847
diabp	-0.027	0.008	-3.341	0.001
totchol	0.005	0.002	3.029	0.002
as.factor(educ)2	0.922	0.284	3.245	0.001
as.factor(educ)3	-0.330	0.326	-1.014	0.311
as.factor(educ)4	-0.481	0.372	-1.293	0.196

```
glmer(cursmoke ~ age + sex + totchol + as.factor(educ) + bmi + diabetes + heartrte + prevchd + prevstrk +
```

What to Turn In: For this assignment you will turn in a single pdf document with (a) your summary of findings and (b) an Appendix with figures and (c) an Appendix with all R code (in the form of a knited pdf).

## OBJECTIVE:

An objective or description of the goals of the analysis

We are interested to describe the smoking habits of the participants in the Framingham Heart study as they age and the impact of smoking on certain health outcomes. The Framingham heart study asks participants about their smoking habits at each visit. In particular, participants are asked if they are currently smoking at this visit (0 = Not a current smoker, 1 = Current smoker), which we will refer to as current smoking status. In addition, participants also report the number of cigarettes they are smoking per day. A more complete description of each of variables in the Framingham Heart study can be found in the Framingham Heart Study Longitudinal Data Documentation.

We are interested to answer the following questions:

- (1) Is there a relationship between age and smoking status? Does this relationship differ by sex?
- (2) Is there a relationship between the number of cigarettes smoked per day and age? Does this relationship differ by sex?
- (3) The relationship between current smoking status and systolic blood pressure.
- (4) The relationship between current smoking status and diastolic blood pressure.
- (5) The relationship between current smoking status and serum total cholesterol.

## STUDY DESIGN:

A brief description of the study design and the data

## **METHODS:**

A methods section describing your statistical analysis (please justify all modeling choices that were made with evidence).

## **RESULTS:**

A results section that includes a) descriptive statistics for the data b) a summary of your key findings including supporting numerical summaries (i.e. confidence intervals, pvalues, etc.) c) interpretations of your key findings (i.e. interpretations of coefficients).

## **CONCLUSION:**

A conclusion specifically answering the objective of the analysis.

## **APPENDIX:**