

Final Project

Volha Tryputsen

12/11/2018

Contents

EDA	4
response vs continuous covariates	4
(1): Smoking ~ age, sex	4
(2): Smoking ~ totchol, sex	6
(3): Smoking ~ sysbp, sex	8
(4): smoking ~ diabp, sex	9
(5): smoking ~ cigpday, sex	10
(6): smoking ~ bmi, sex	11
(7): smoking ~ heartrte, sex	13
(8): smoking ~ glucose, sex	14
(9): smoking ~ timeap, sex	15
(10): smoking ~ timemi, sex [not sig]	16
(11): smoking ~ timemifc, sex [not sig]	17
(12): smoking ~ timemchd, sex [not sig]	18
(13): smoking ~ timestrk, sex [not sig]	19
(14): smoking ~ timecvd, sex [not sig]	20
(14): smoking ~ timecvd, sex [possibly?]	21
(15): smoking ~ timedth, sex [possibly?]	22
(16): smoking ~ timehyp, sex [possibly?]	23
response vs categorical covariates	24
smoking ~ sex	24
smoking ~ diabetes	24
smoking ~ bpmeds	25
smoking ~ educ	25
smoking ~ prevchd	25
smoking ~ prevap	25
smoking ~ prevmi [not sig]	25
smoking ~ prevstrk [sig all]	25
smoking ~ prevhyp [1/2 sig]	25
smoking ~ death [1/2 sig]	26
smoking ~ death [1/2 sig]	26
smoking ~ angina [1/2 sig]	26
smoking ~ mi_fchd	26
smoking ~ anychd [not sig]	26
smoking ~ stroke [not sig]	26
smoking ~ cvd [borderline sig]	26
smoking ~ hyperten [highly sig]	27
pairwise correlations	27
demographics	28
education	28
Model	29
Smoking ~ age+ sex + covariates	29
demographics	30
heart	32

blood pressure - nothing is significant	35
diabetes - nothing	43
Prevalence	47
Number of cigarettes ~ age + sex + covariates	55
systolic bp ~ smoking+ covariates	55
diastolic bp ~ smoking + covariates	55
total cholesterol ~ smoking + covariates	55

```

library(tidyverse)
library(janitor)
library(knitr)
library(gee)
library(MuMIn)
library(forcats)
library(knitr)
library(lme4)

dat =
  read.csv(file = '../.../frmgham2.csv') %>%
  clean_names()

dim(dat)

glimpse(dat)
head(dat)
View(dat)

dat %>% select(sex) %>% head()

apply(dat,2,function(x)100*sum(is.na(x))/11627)

##      randid      sex    totchol      age     sysbp     diabp
## 0.00000000 0.00000000 3.51767438 0.00000000 0.00000000 0.00000000
##  cursmoke   cigpday      bmi   diabetes    bpmeds  heartrte
## 0.00000000 0.67945300 0.44723488 0.00000000 5.10019782 0.05160403
##   glucose     educ   prevchd    prevap   prevmi  prevstrk
## 12.38496603 2.53719790 0.00000000 0.00000000 0.00000000 0.00000000
##   prevhyp     time    period     hdlc     ldlc     death
## 0.00000000 0.00000000 0.00000000 73.96576933 73.97437000 0.00000000
##    angina    hospmi    mi_fchd   anychd     stroke     cvd
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   hyperten   timeap   timemi   timemifc  timechd  timestrk
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##    timecvd   timedth   timehyp
## 0.00000000 0.00000000 0.00000000

dat =
  dat %>%
  mutate(sex = factor(sex)) #>%>% fct_recode(sex, "female" = "2", "male" = "1")

# what variables have more than 70% data missing
apply(dat,2,function(x)100*sum(is.na(x))/11627)[apply(dat,2,function(x)100*sum(is.na(x))/11627) > 70]

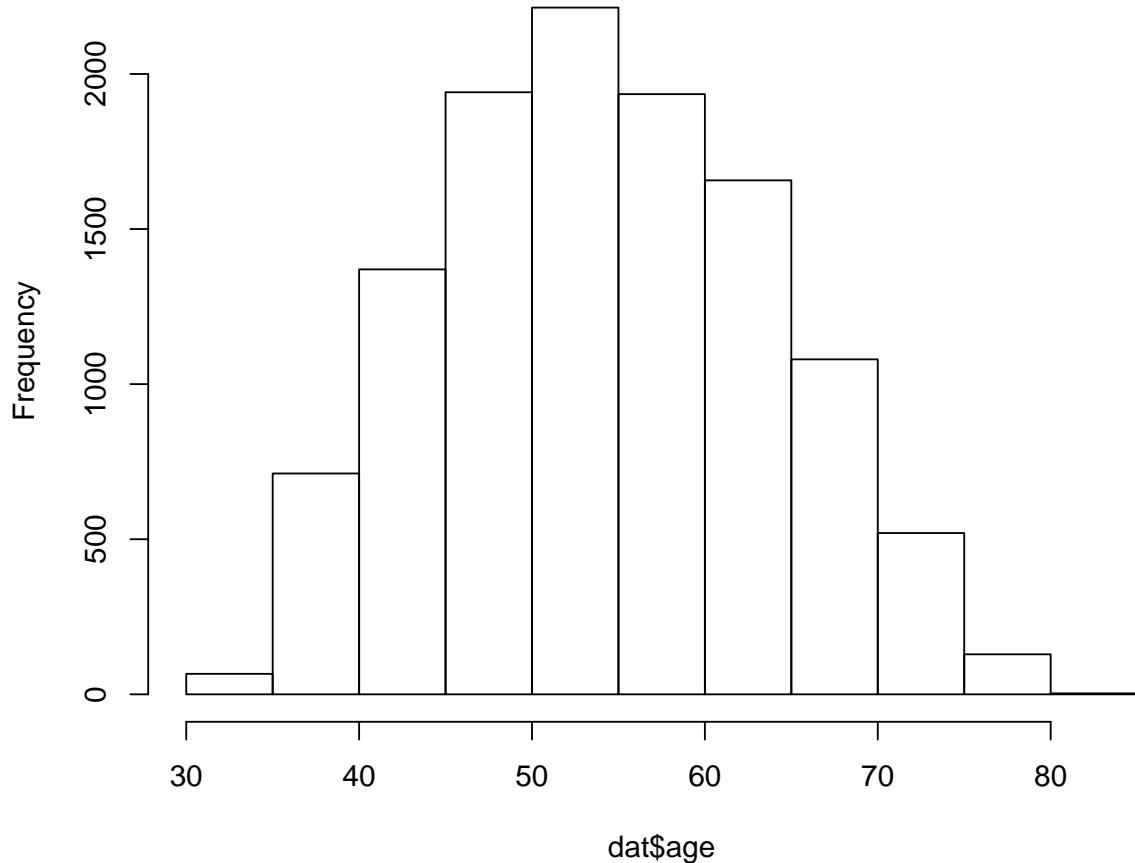
##      hdlc      ldlc
## 73.96577 73.97437

```

```
# exclude those variables( hdlc, ldlc)
dat = dat[,-c(22, 23)]
```

```
#center age for interpretation
hist(dat$age)
```

Histogram of dat\$age



```
mean(dat$age)
```

```
## [1] 54.79281
```

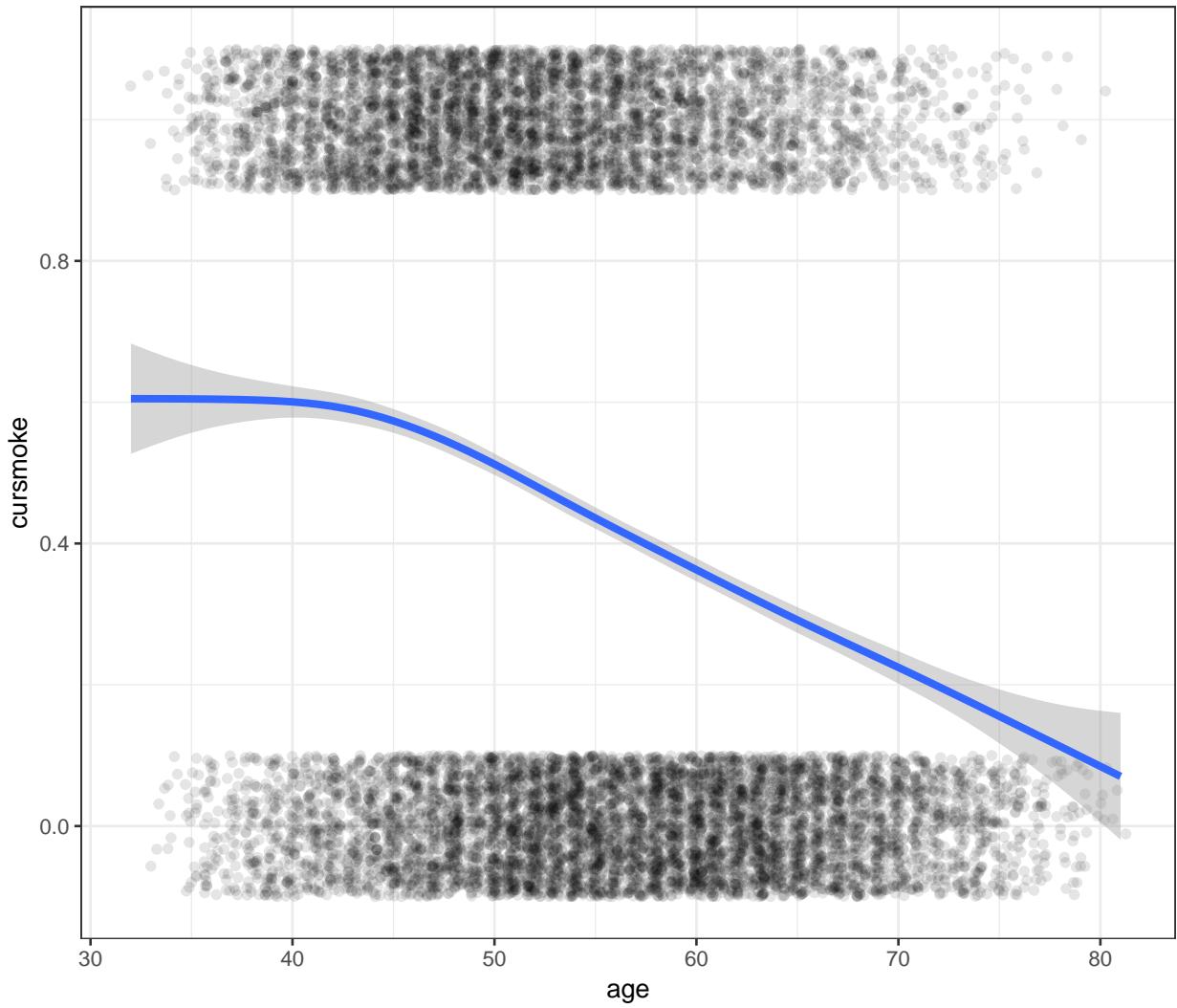
```
dat = dat %>% mutate(cage = age - median(dat$age[dat$period == "1"]))
```

EDA

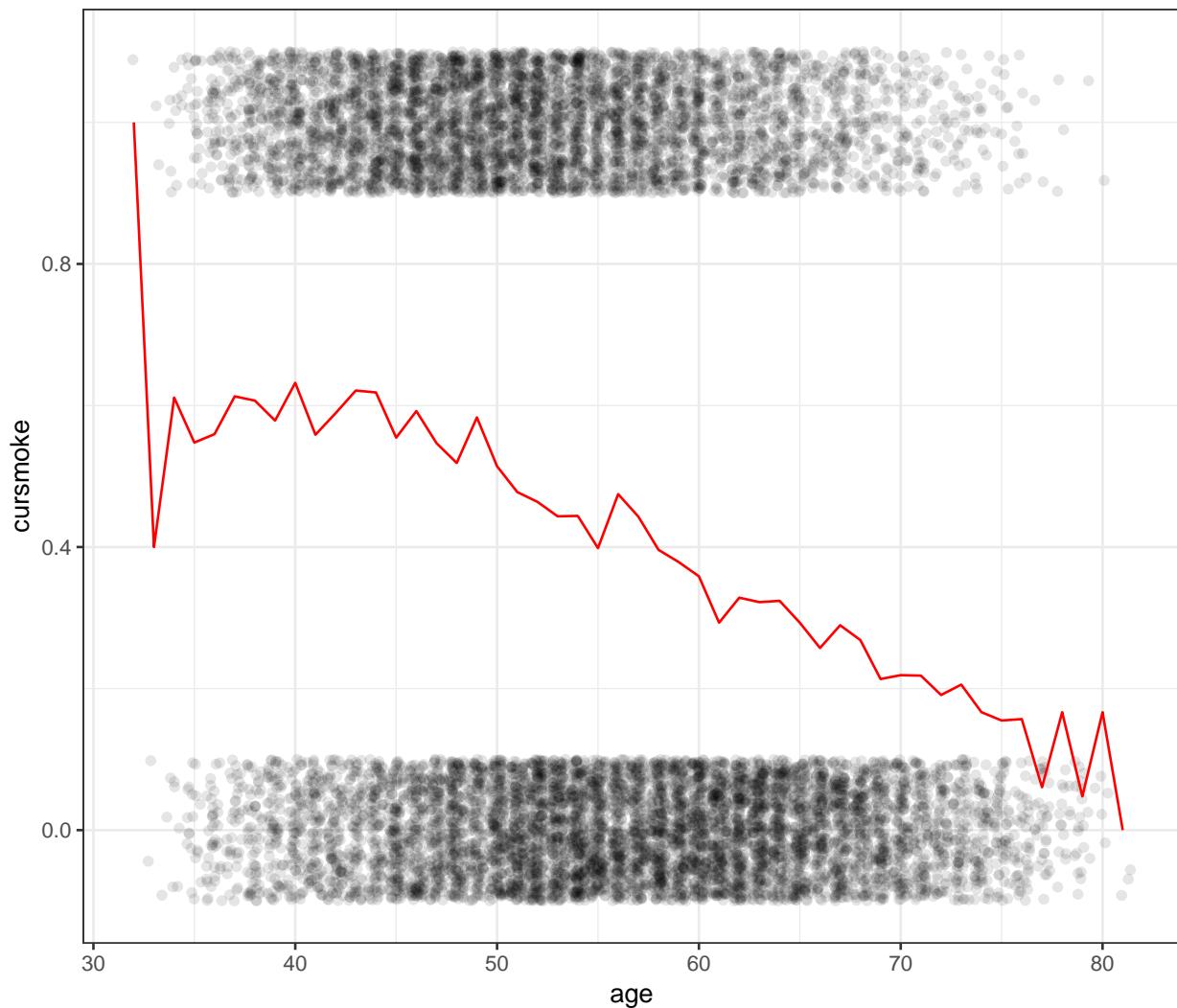
response vs continuous covariates

(1): Smoking ~ age, sex

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



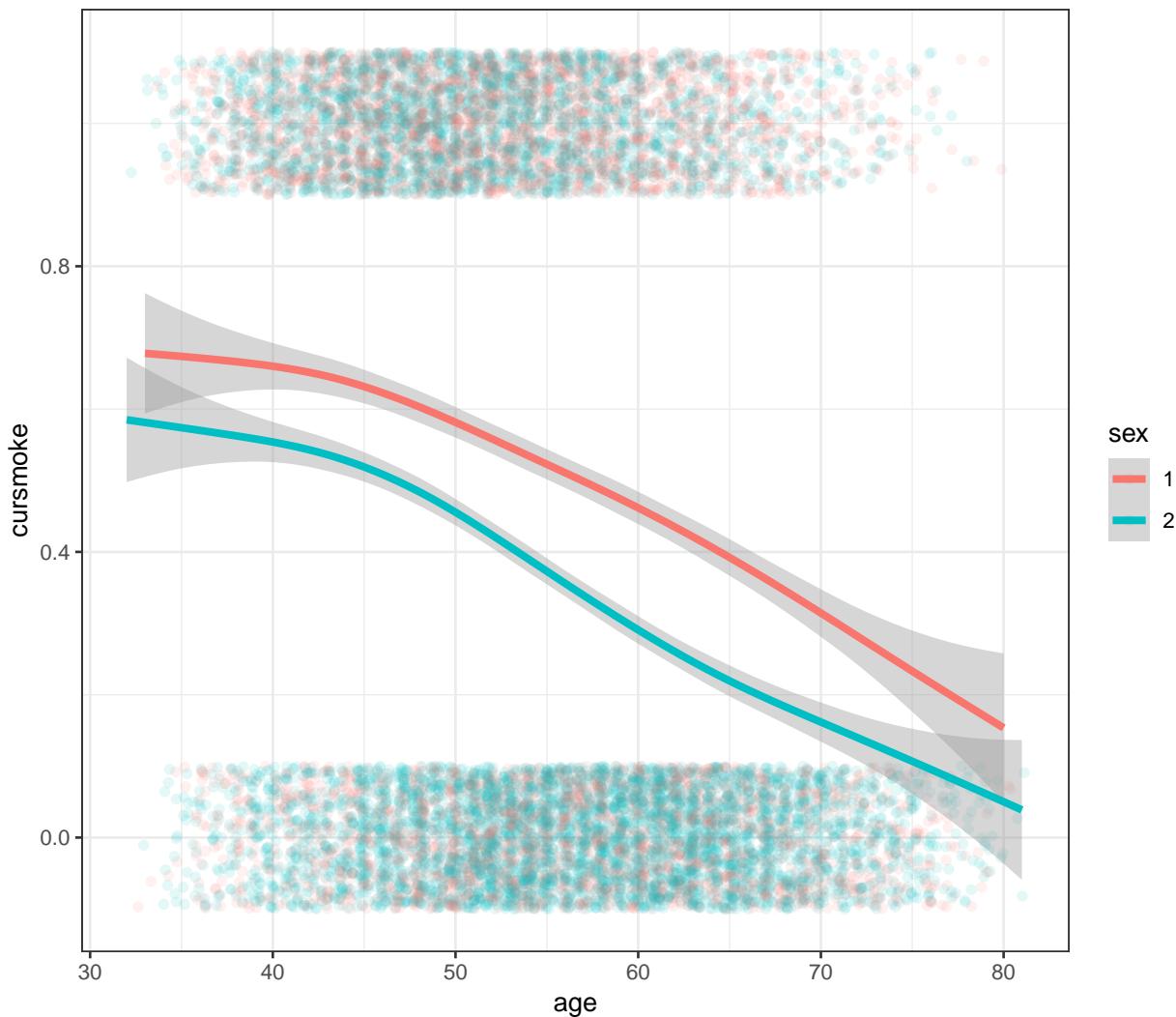
```
dat %>%
  select(cursmoke, age) %>%
  ggplot(aes(x = age, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  stat_summary(fun.y = 'mean', geom="line", col = 'red') +
  theme_bw()
```



The proportion of smokers decreases with the age.

Smoking ~ age, sex

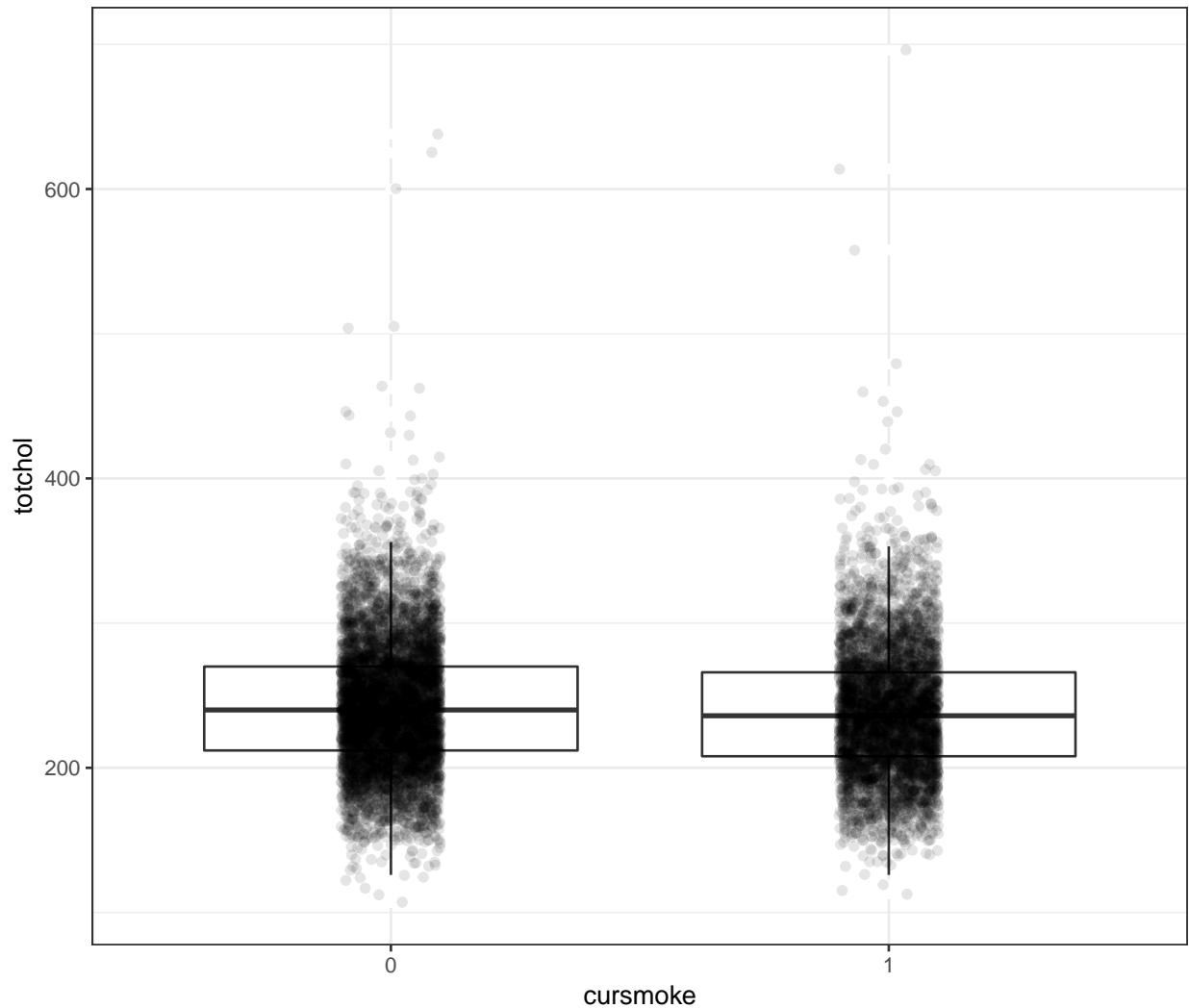
```
dat %>%
  select(cursmoke, age, sex) %>%
  ggplot(aes(x = age, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```



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

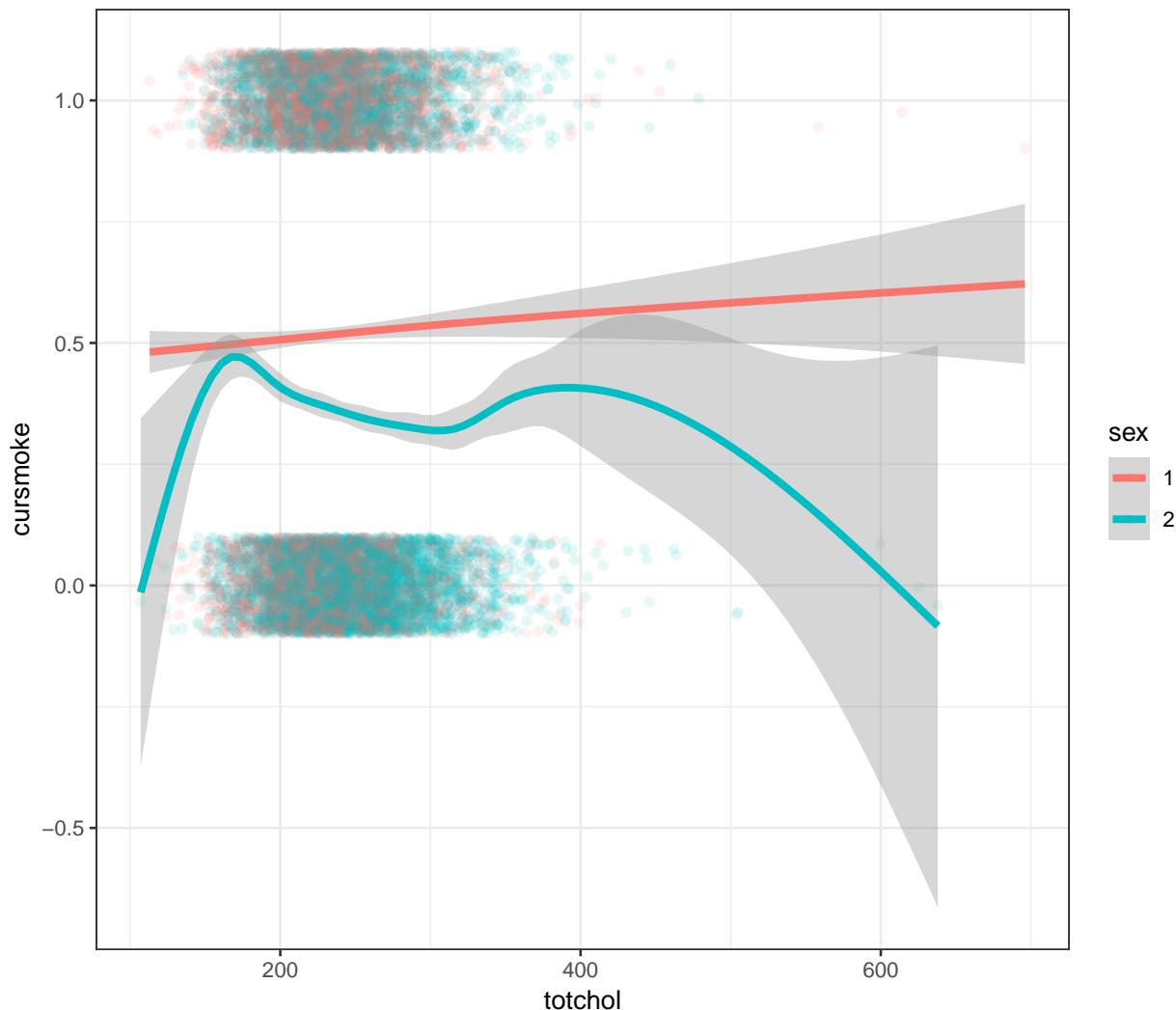
(2): Smoking ~ totchol, sex

```
dat %>%
  select(cursmoke, totchol) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = totchol, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  geom_jitter(width = 0.1, alpha = 0.1) +
  theme_bw()
```



smoking ~ totchol x sex [interaction]

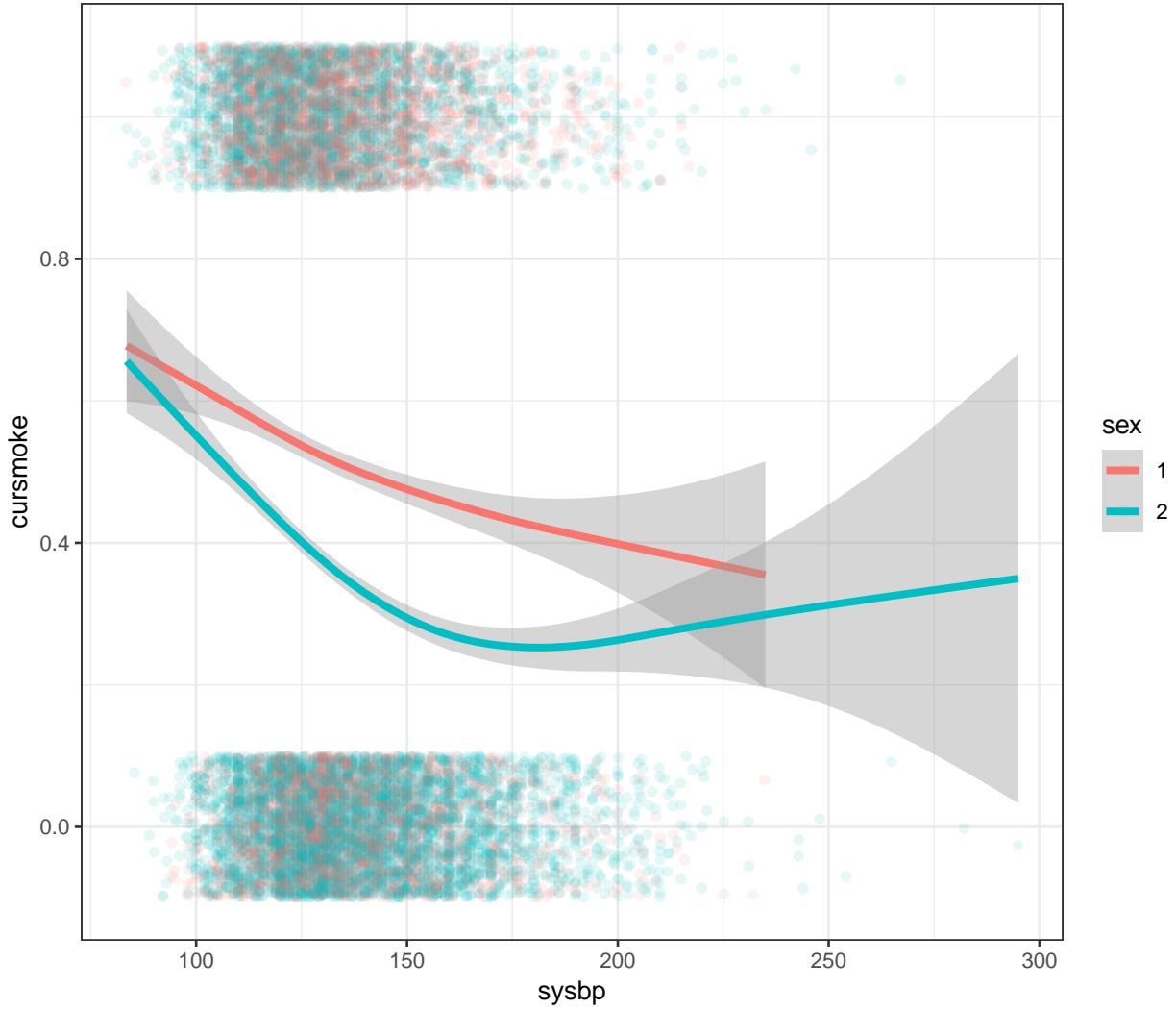
```
dat %>%
  select(cursmoke, totchol, sex) %>%
  ggplot(aes(x = totchol, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```



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).

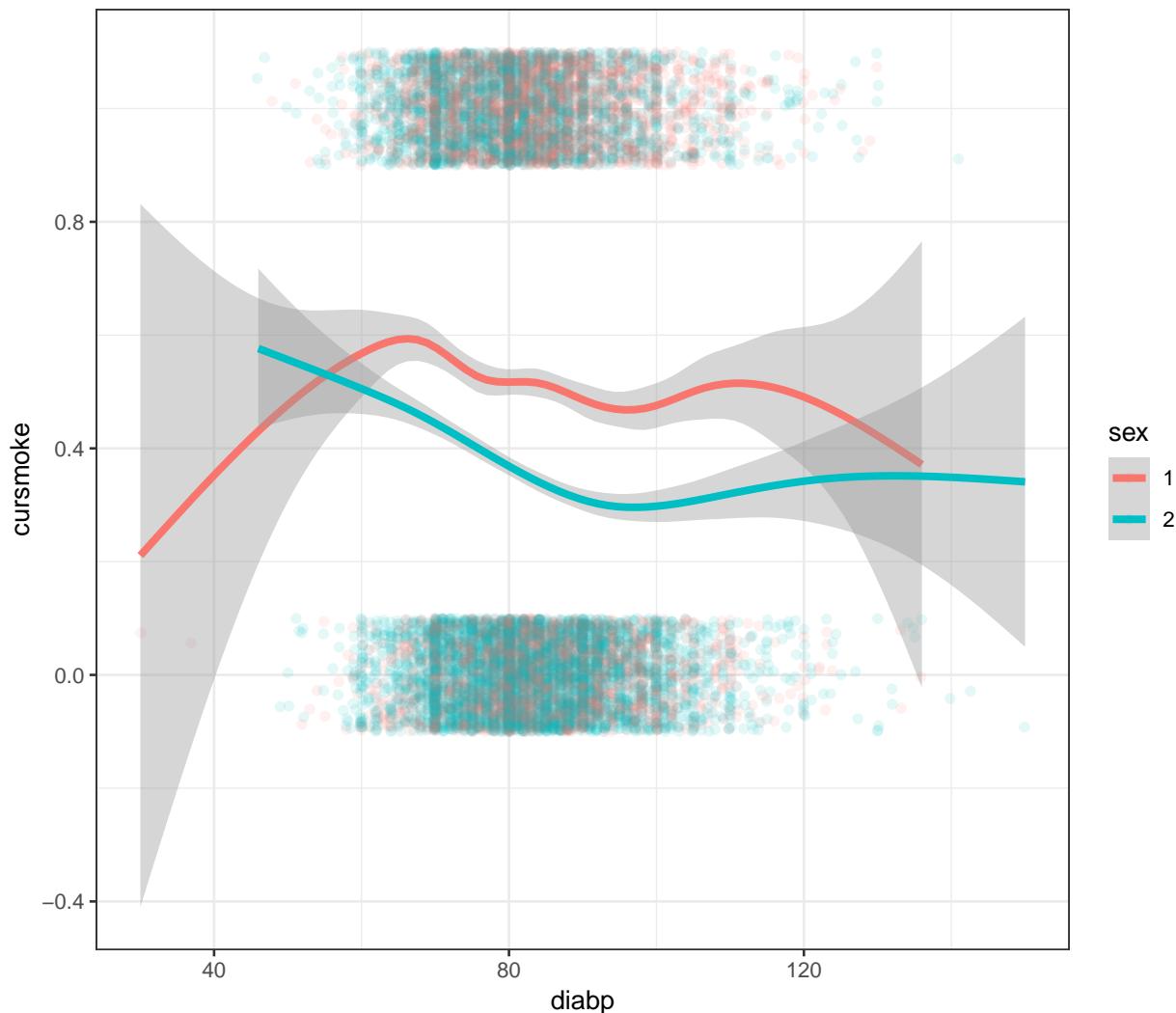
(3): Smoking ~ sysbp, sex

```
dat %>%
  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()
```



(4): smoking ~ diabp, sex

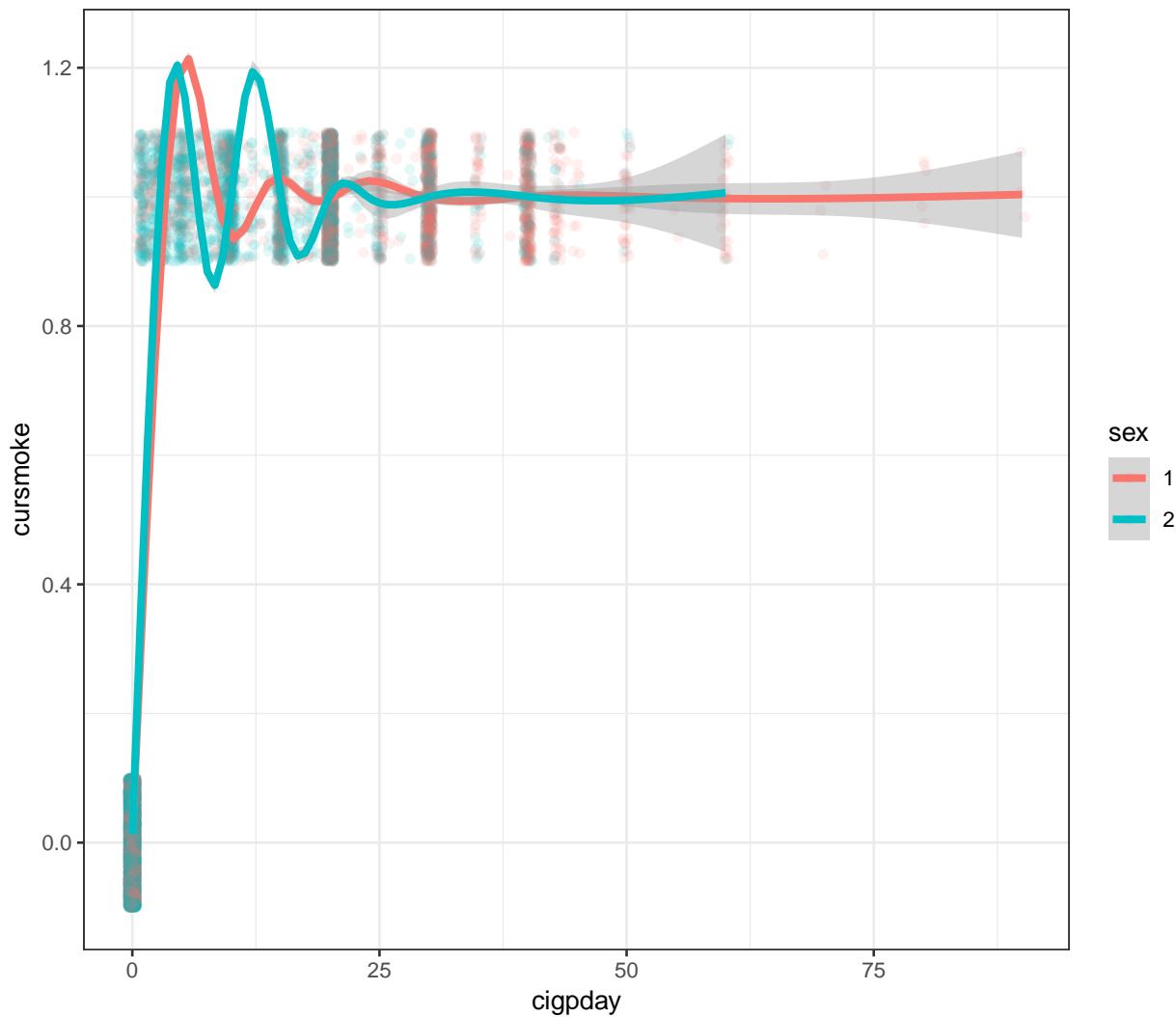
```
dat %>%
  select(cursmoke, diabp, sex) %>%
  ggplot(aes(x = diabp, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```



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

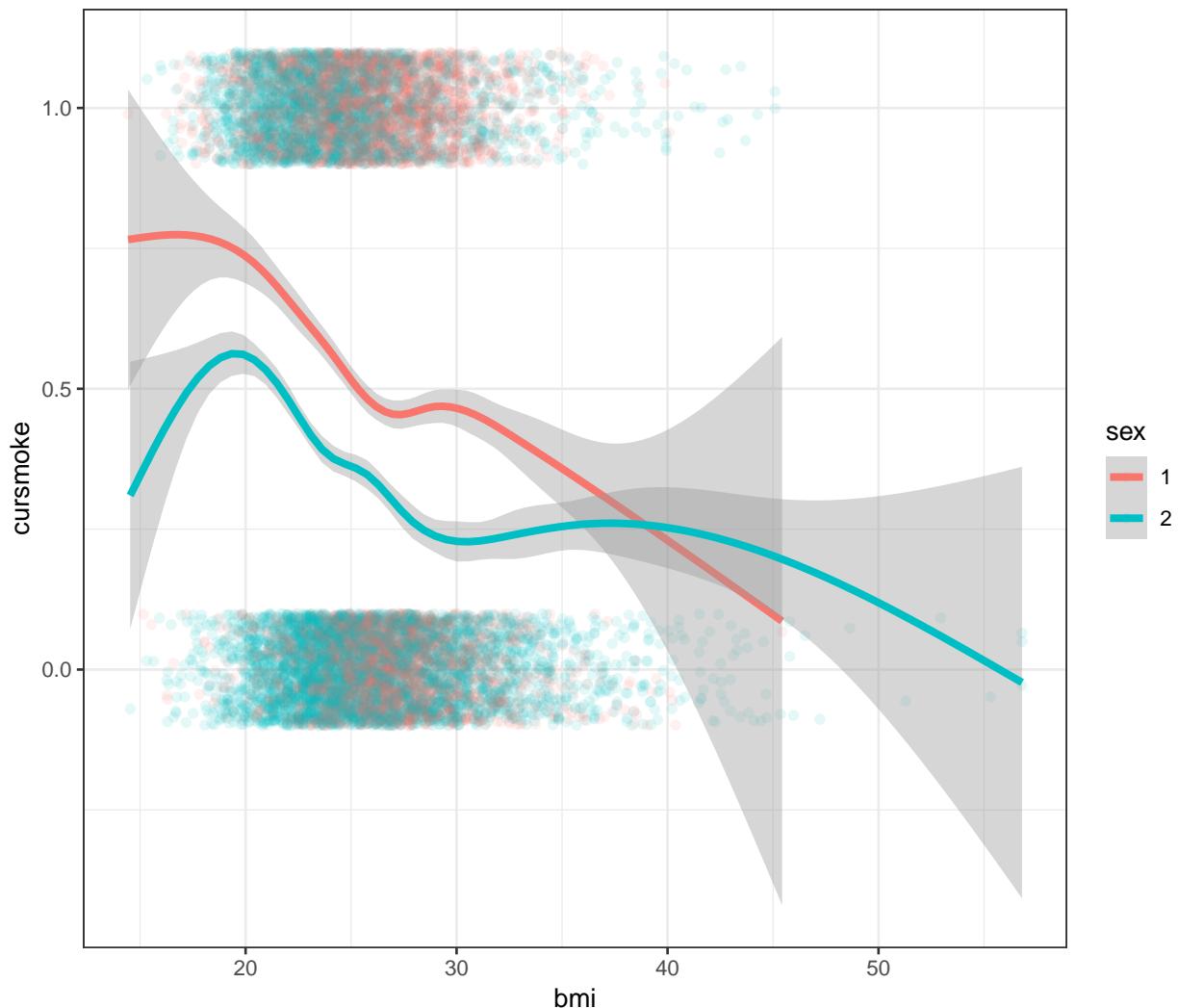
(5): smoking ~ cigpday, sex

```
dat %>%
  select(cursmoke, cigpday, sex) %>%
  ggplot(aes(x = cigpday, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  #stat_summary(fun.y = 'mean', geom="line", col = 'red') +
  theme_bw()
```

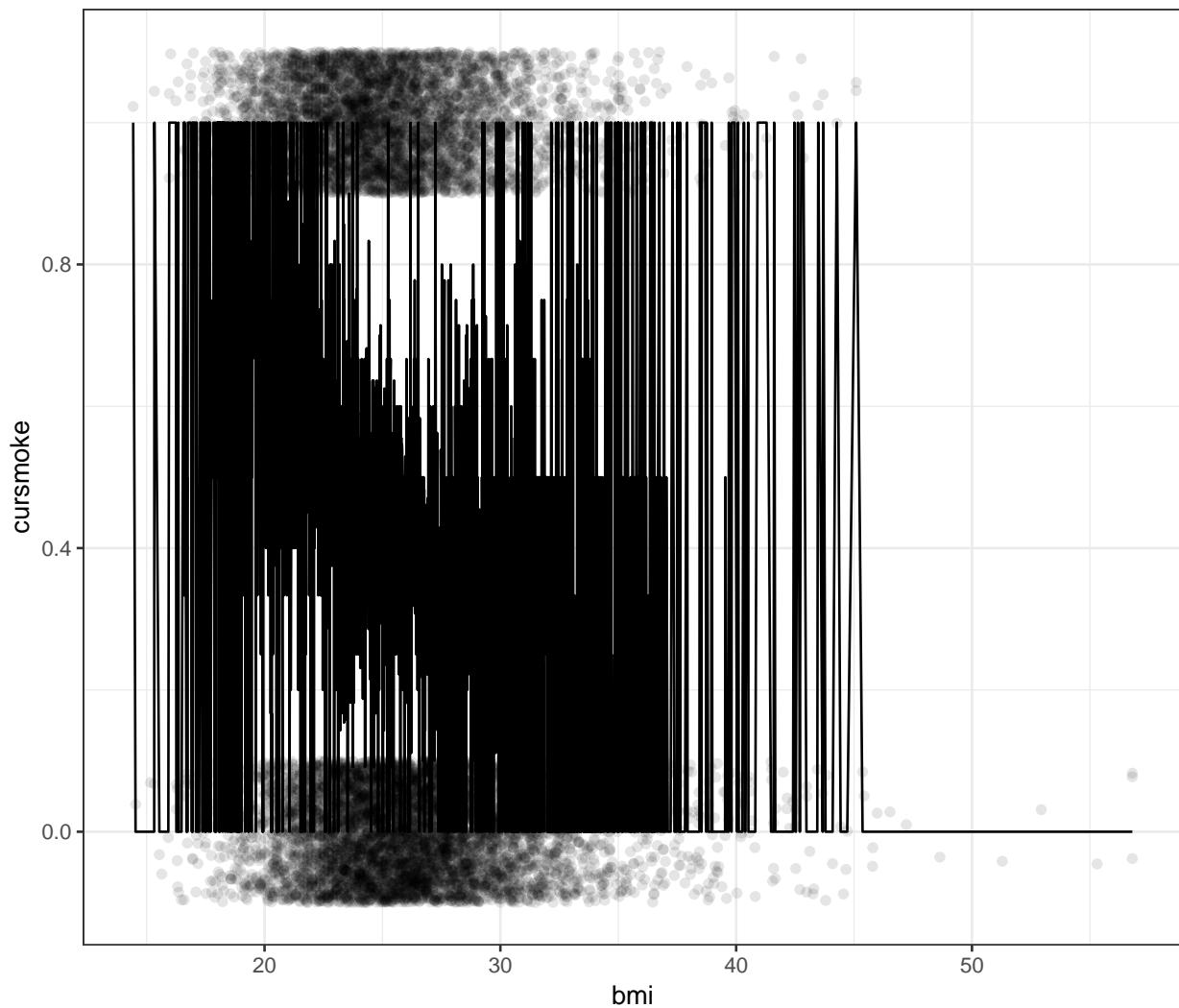


(6): smoking ~ bmi, sex

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

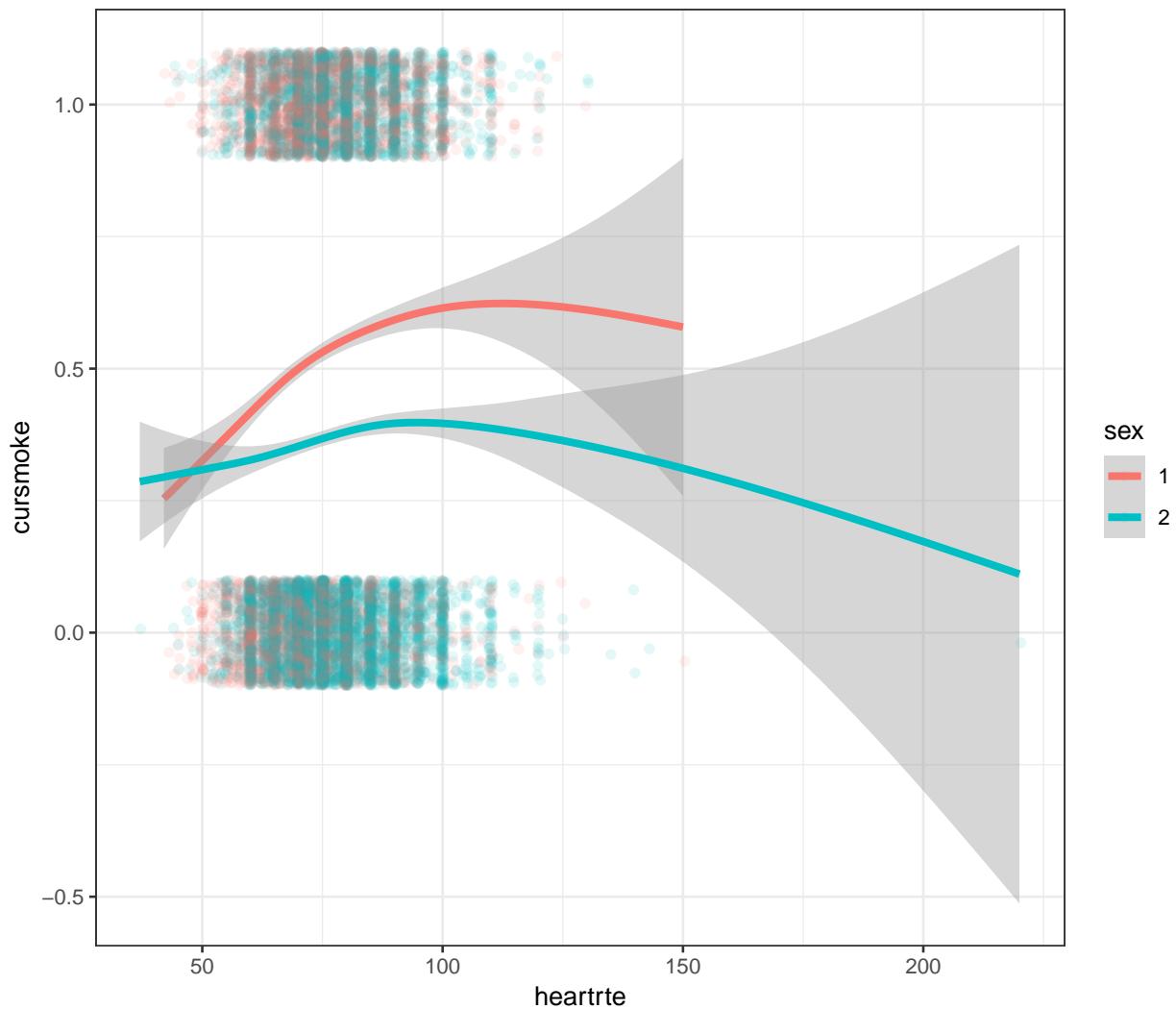


```
dat %>%
  select(cursmoke, bmi) %>%
  ggplot(aes(x = bmi, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  stat_summary(fun.y = 'mean', geom="line") +
  theme_bw()
```



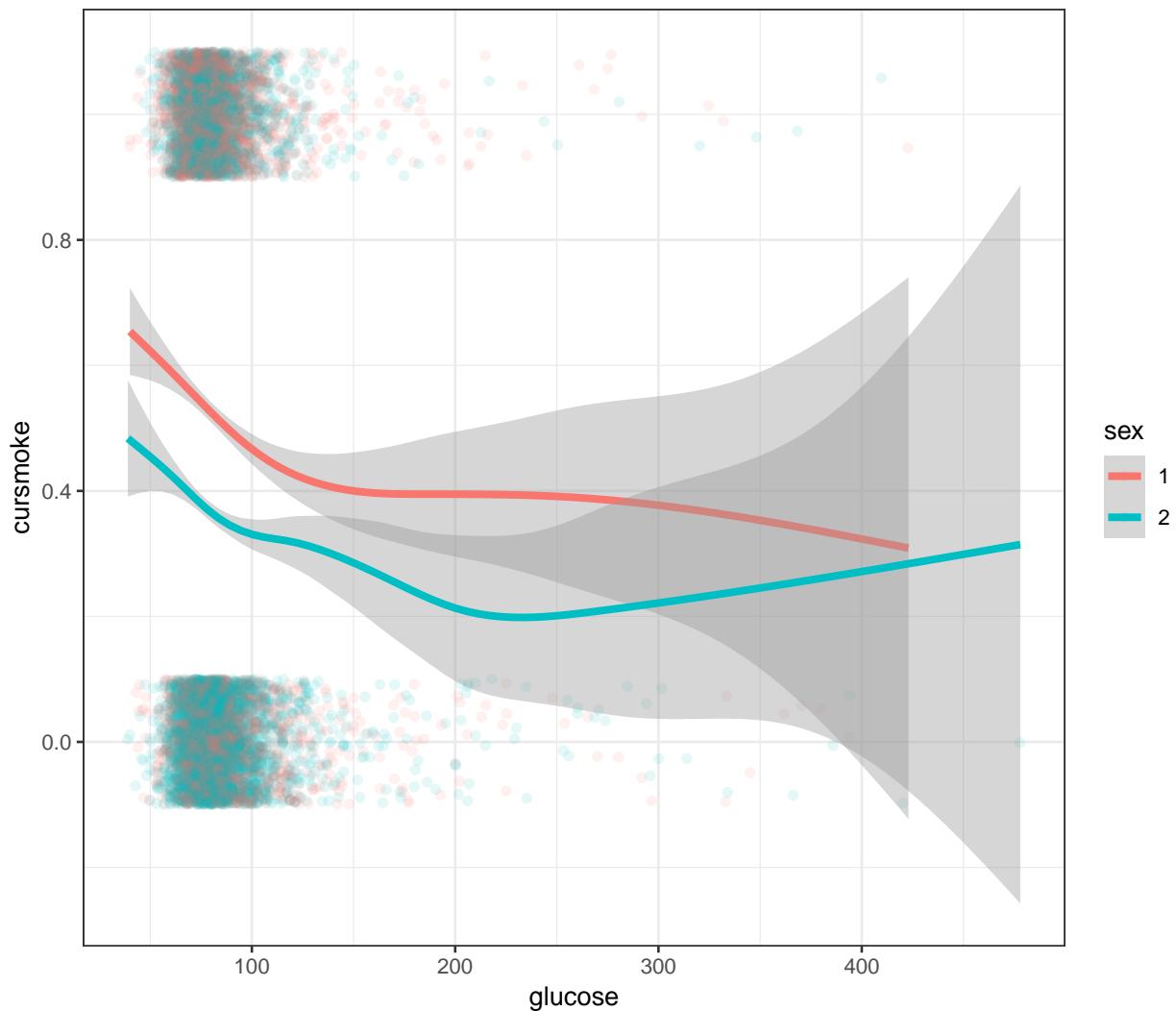
(7): smoking ~ heartrte, sex

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



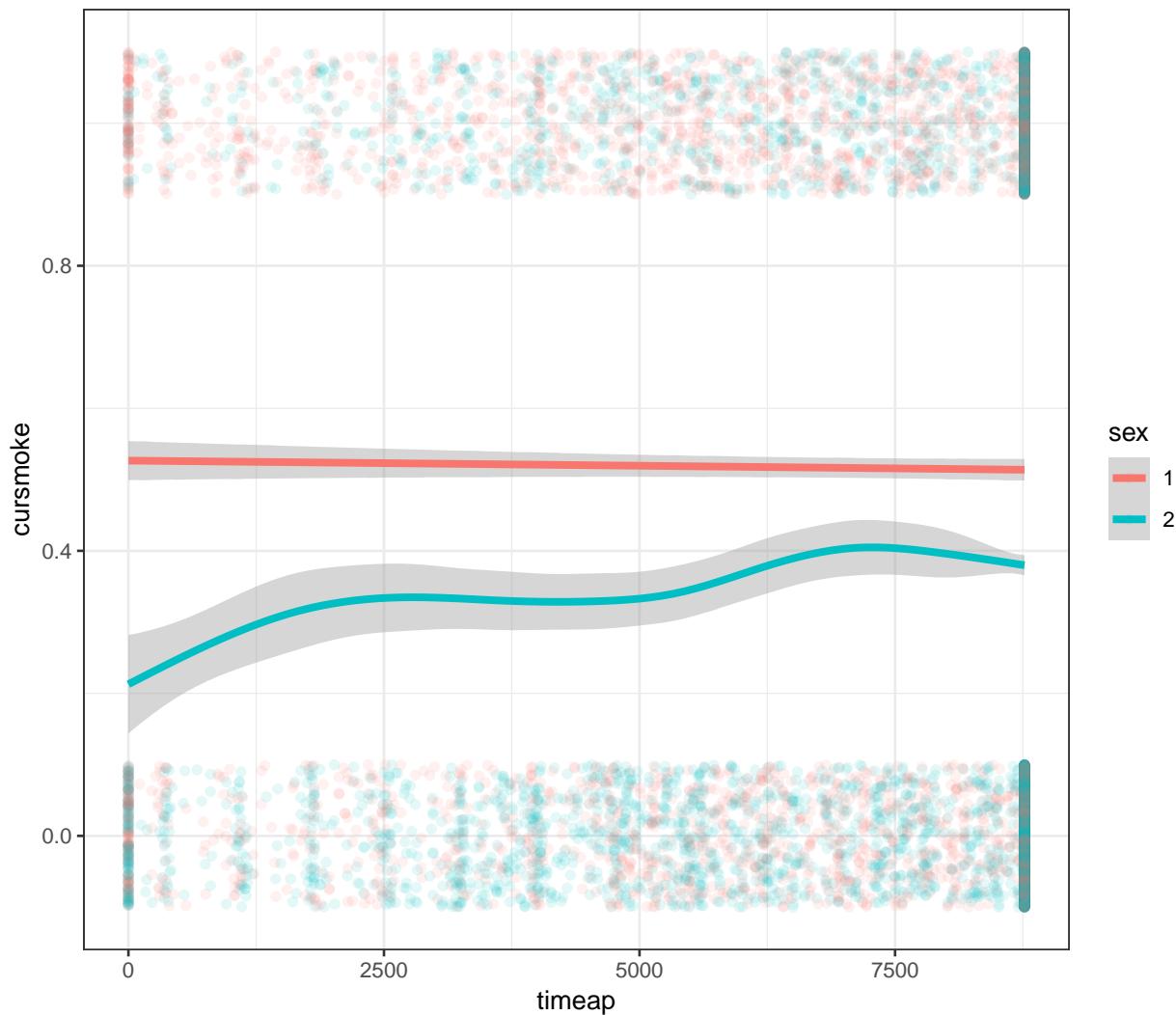
(8): smoking ~ glucose, sex

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



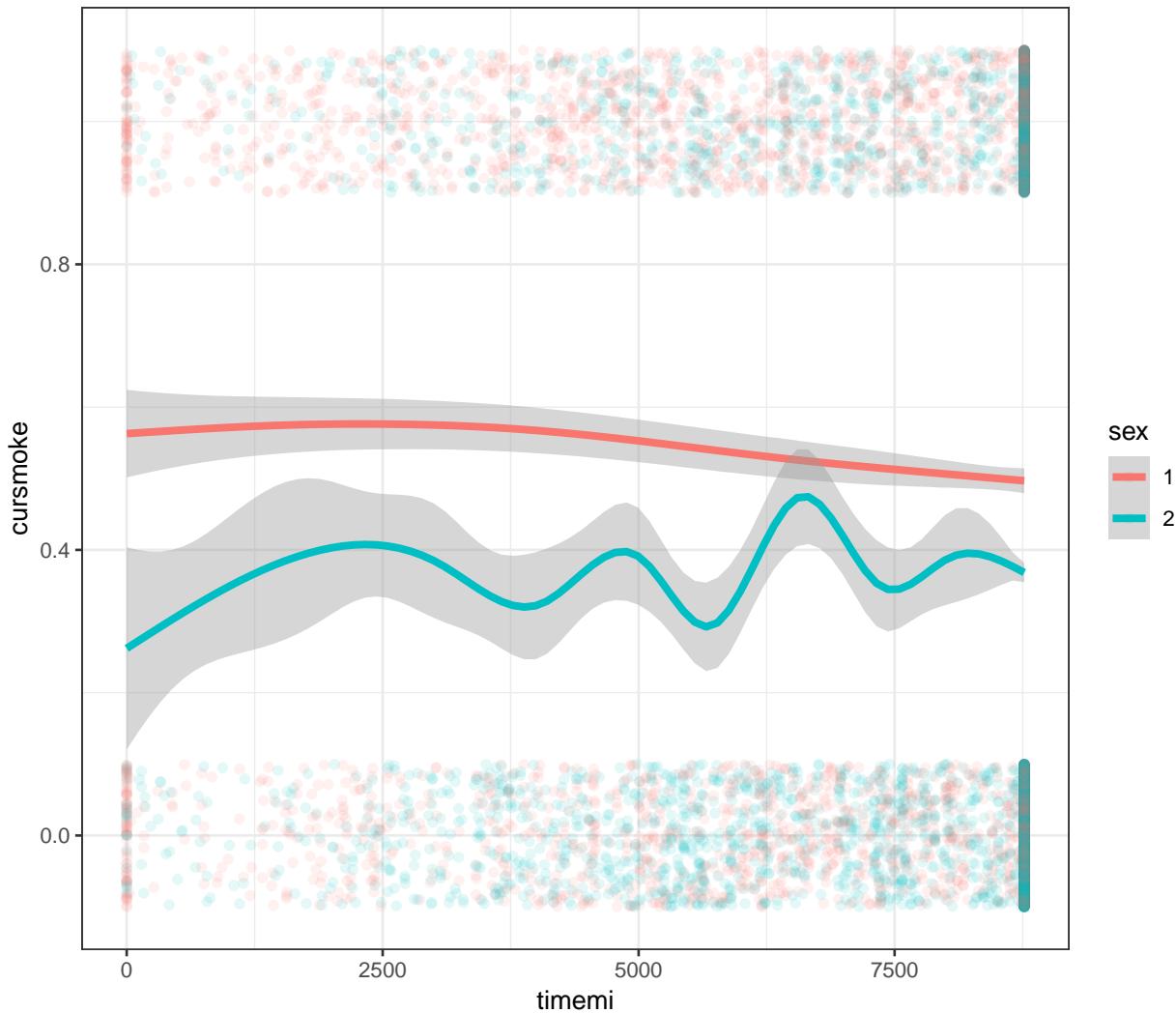
(9): smoking ~ timeap, sex

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



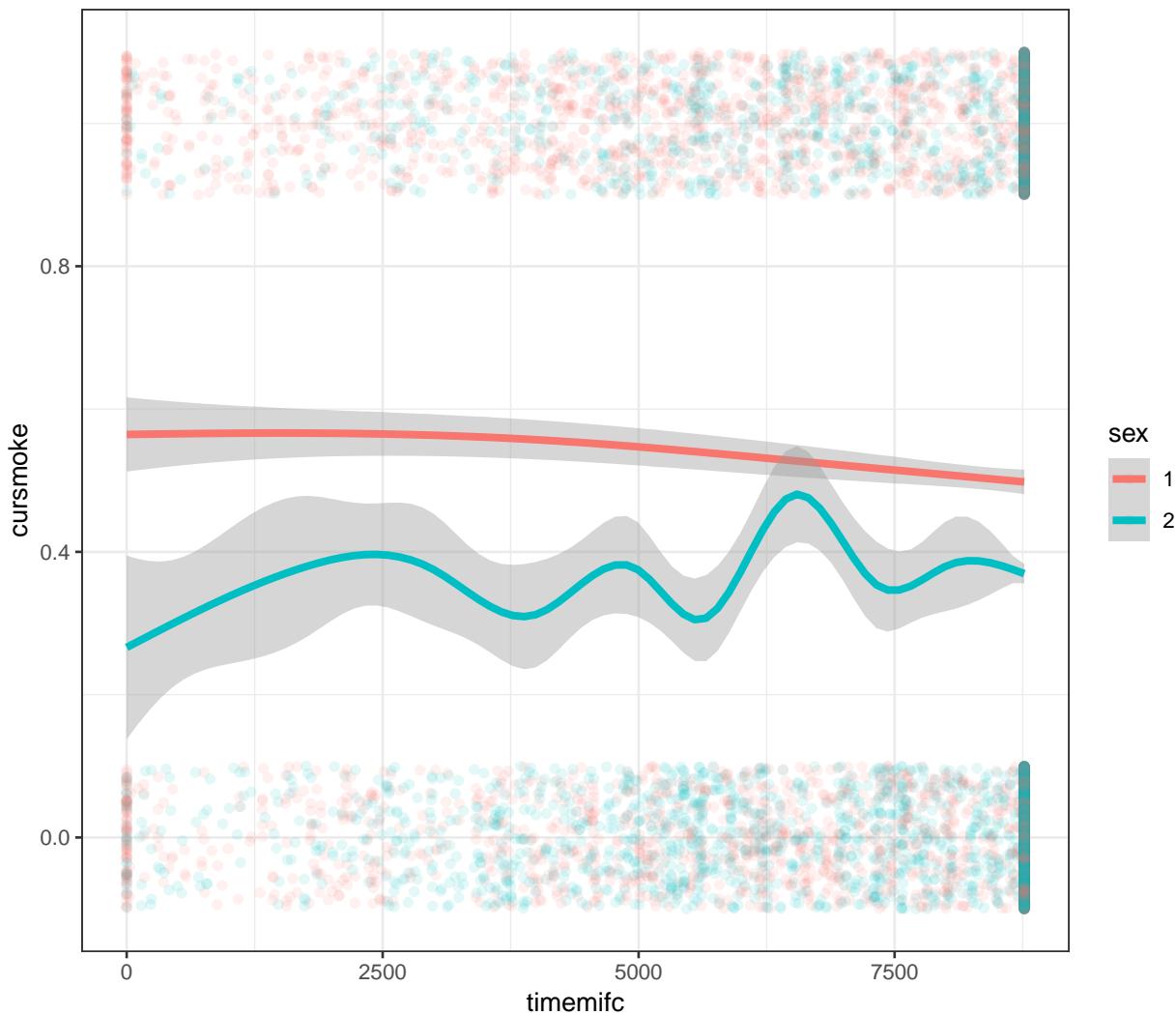
(10): smoking ~ timemi, sex [not sig]

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



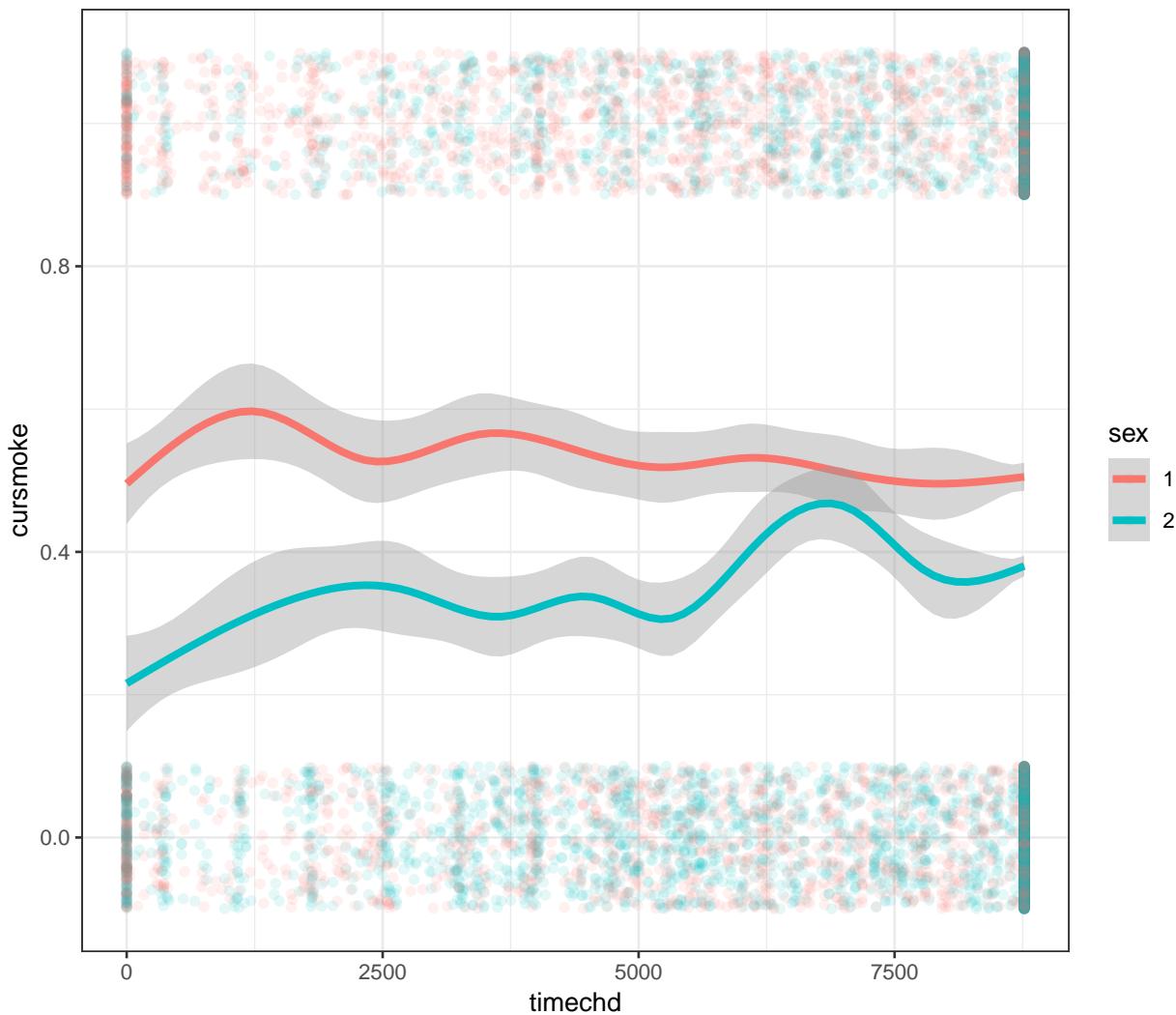
(11): smoking ~ timemifc, sex [not sig]

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



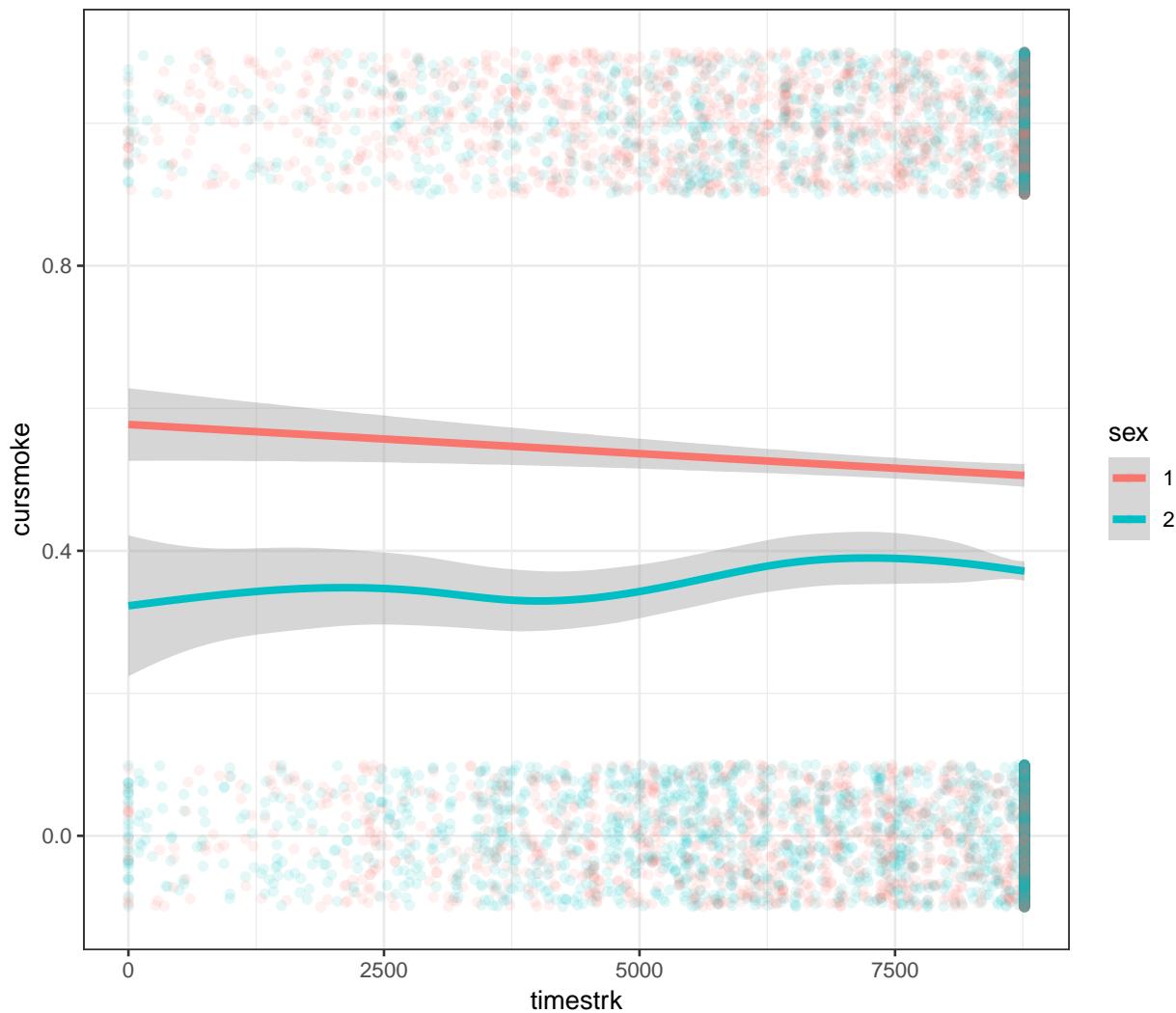
(12): smoking ~ timemchd, sex [not sig]

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



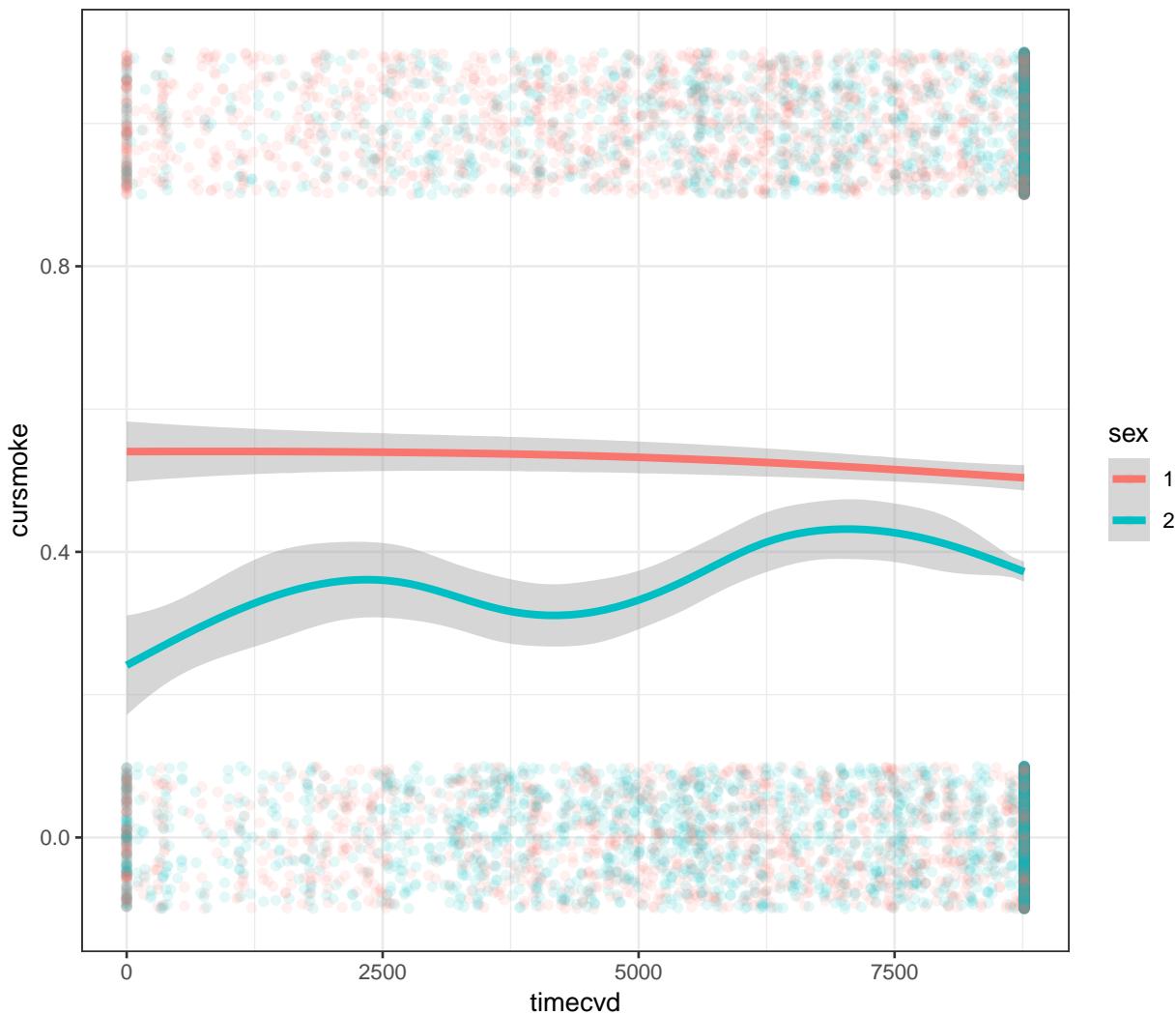
(13): smoking ~ timestrk, sex [not sig]

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



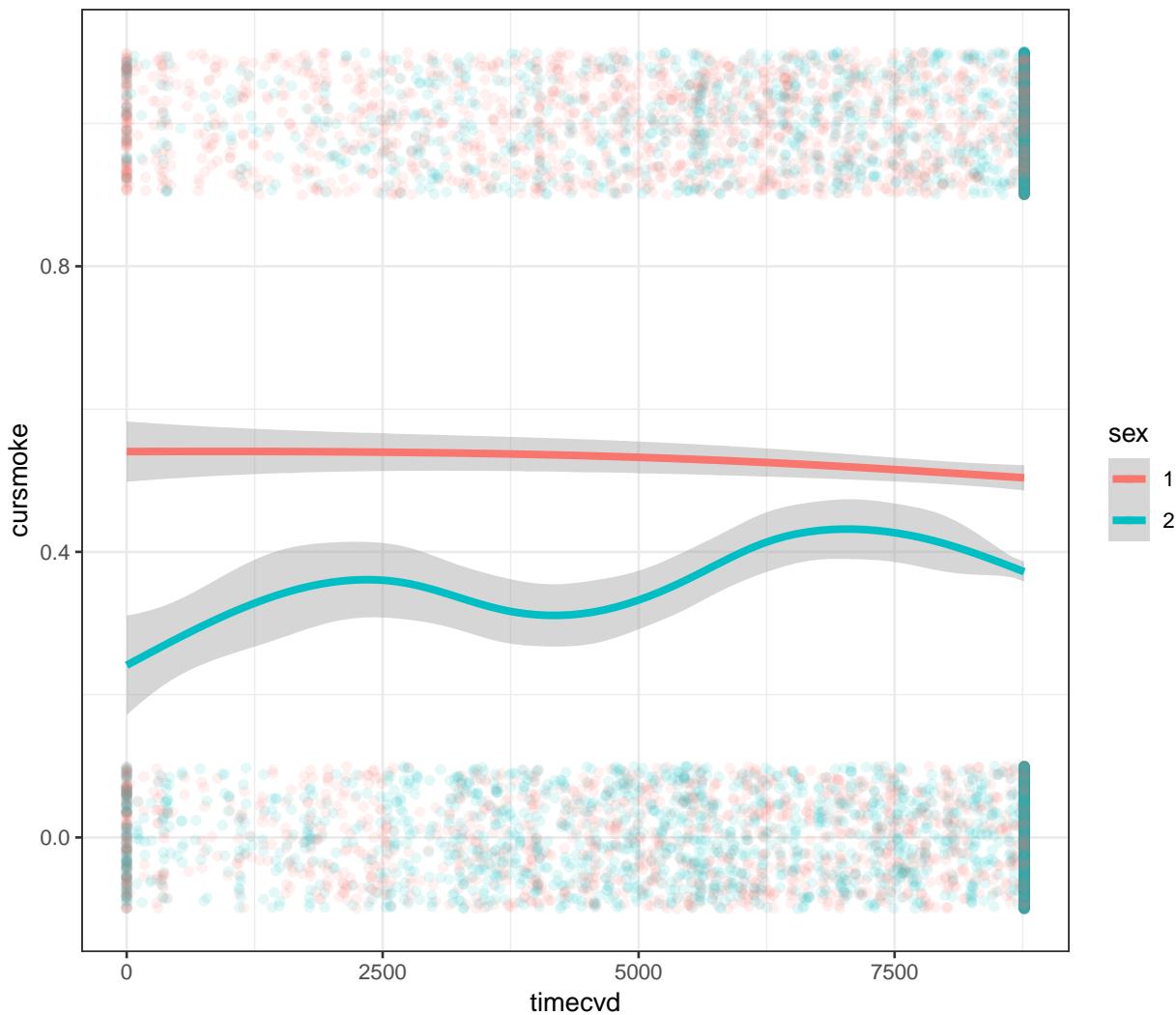
(14): smoking ~ timecvd, sex [not sig]

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



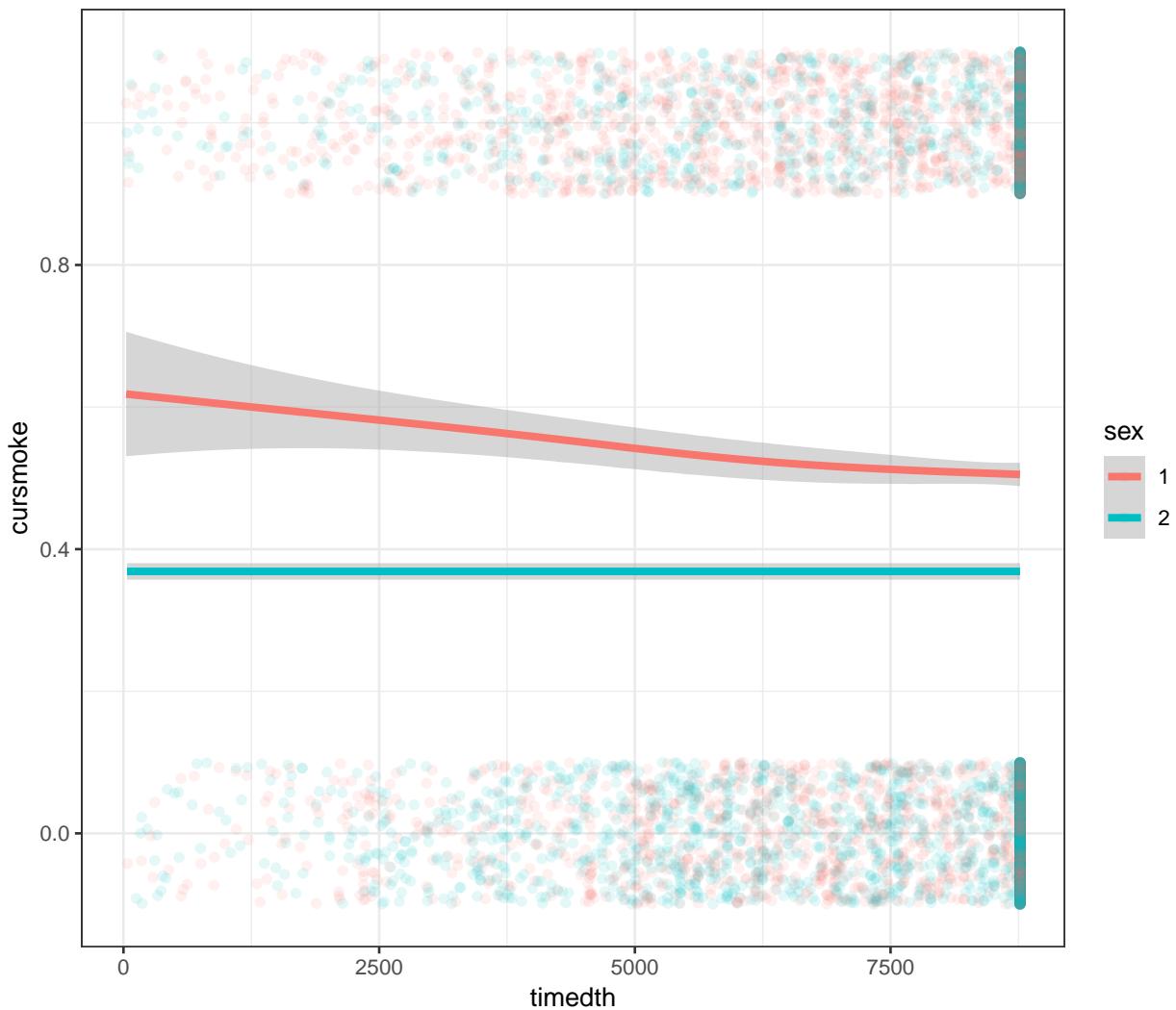
(14): smoking ~ timecvd, sex [possibly?]

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



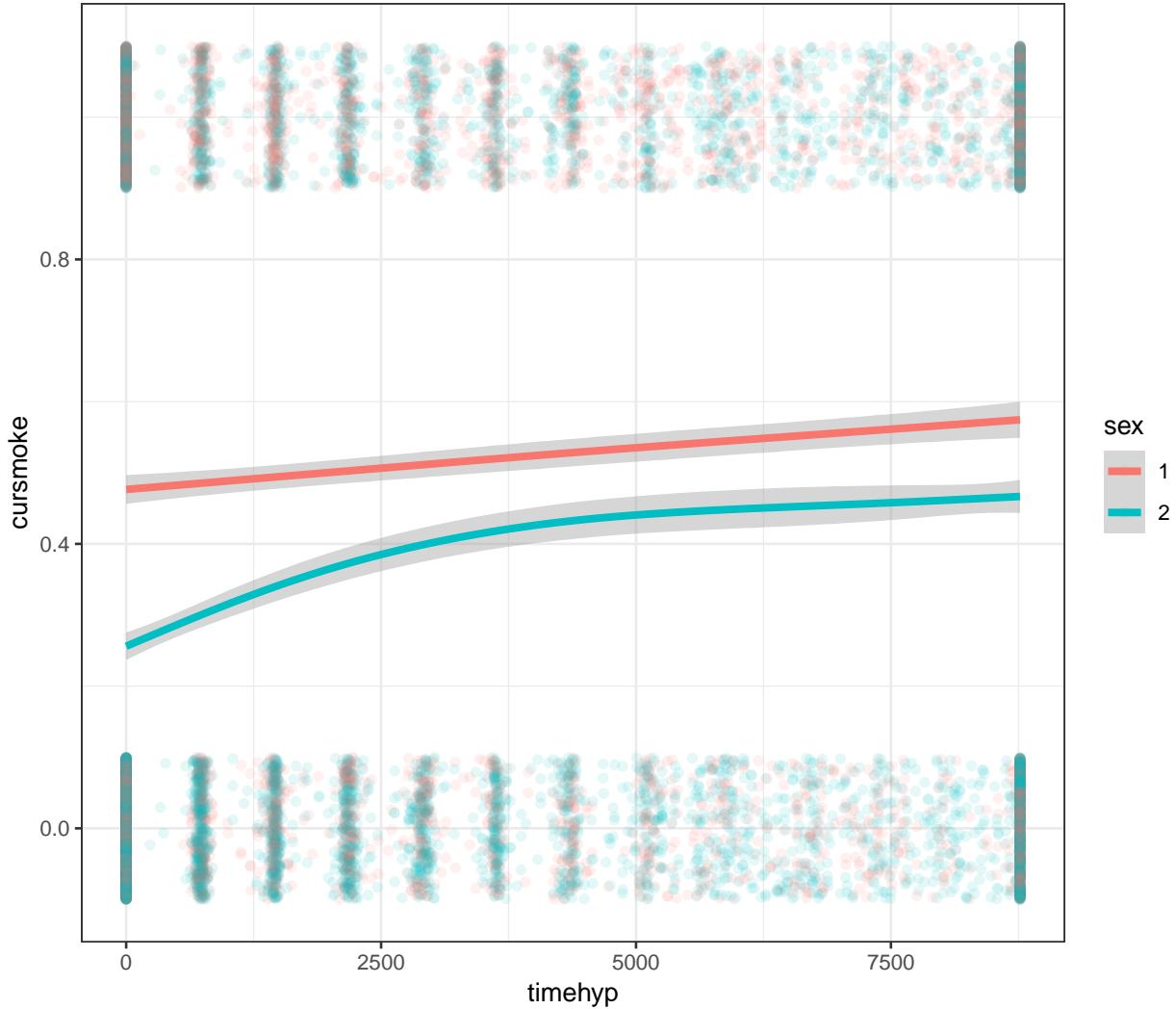
(15): smoking ~ timedth, sex [possibly?]

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



(16): smoking ~ timehyp, sex [possibly?]

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



response vs categorical covariates

smoking ~ sex

```
c(chisq.test(table(dat$sex[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$sex, dat$cursmoke))$p.value)
```

```
## [1] 7.214771e-40 4.405111e-57
```

smoking ~ diabetes

```
c(chisq.test(table(dat$diabetes[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$diabetes, dat$cursmoke))$p.value)
```

```
## [1] 1.703237e-03 3.426931e-06
```

smoking ~ bpmeds

```
c(chisq.test(table(dat$bpmeds[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$bpmeds, dat$cursmoke))$p.value)
```

```
## [1] 1.009822e-03 1.967621e-19
```

smoking ~ educ

```
c(chisq.test(table(dat$educ[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$educ, dat$cursmoke))$p.value)
```

```
## [1] 4.204420e-04 4.992107e-13
```

smoking ~ prevchd

```
c(chisq.test(table(dat$prevchd[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$prevchd, dat$cursmoke))$p.value)
```

```
## [1] 1.899578e-01 2.980984e-06
```

smoking ~ prevap

```
c(chisq.test(table(dat$prevap[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$prevap, dat$cursmoke))$p.value)
```

```
## [1] 2.053074e-02 2.037869e-07
```

smoking ~ prevmi [not sig]

```
c(chisq.test(table(dat$prevmi[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$prevmi, dat$cursmoke))$p.value)
```

```
## [1] 0.2575237 0.8932076
```

smoking ~ prevstrk [sig all]

```
c(chisq.test(table(dat$prevstrk[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$prevstrk, dat$cursmoke))$p.value)
```

```
## [1] 0.2501900504 0.0001278106
```

smoking ~ prevhyp [1/2 sig]

```
c(chisq.test(table(dat$prevhyp[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$prevhyp, dat$cursmoke))$p.value)
```

```
## [1] 1.031730e-12 1.449666e-56
```

smoking ~ death [1/2 sig]

```
c(chisq.test(table(dat$death[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$death, dat$cursmoke))$p.value)
```

```
## [1] 0.114047498 0.003851208
```

smoking ~ death [1/2 sig]

```
c(chisq.test(table(dat$death[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$death, dat$cursmoke))$p.value)
```

```
## [1] 0.114047498 0.003851208
```

smoking ~ angina [1/2 sig]

```
c(chisq.test(table(dat$angina[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$angina, dat$cursmoke))$p.value)
```

```
## [1] 0.0724668177 0.0001424006
```

smoking ~ mi_fchd

```
c(chisq.test(table(dat$mi_fchd[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$mi_fchd, dat$cursmoke))$p.value)
```

```
## [1] 1.621399e-03 2.251818e-06
```

smoking ~ anychd [not sig]

```
c(chisq.test(table(dat$anychd[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$anychd, dat$cursmoke))$p.value)
```

```
## [1] 0.6602564 0.9522587
```

smoking ~ stroke [not sig]

```
c(chisq.test(table(dat$stroke[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$stroke, dat$cursmoke))$p.value)
```

```
## [1] 0.6329558 0.3488428
```

smoking ~ cvd [borderline sig]

```
c(chisq.test(table(dat$cvd[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$cvd, dat$cursmoke))$p.value)
```

```
## [1] 0.02670128 0.01432159
```

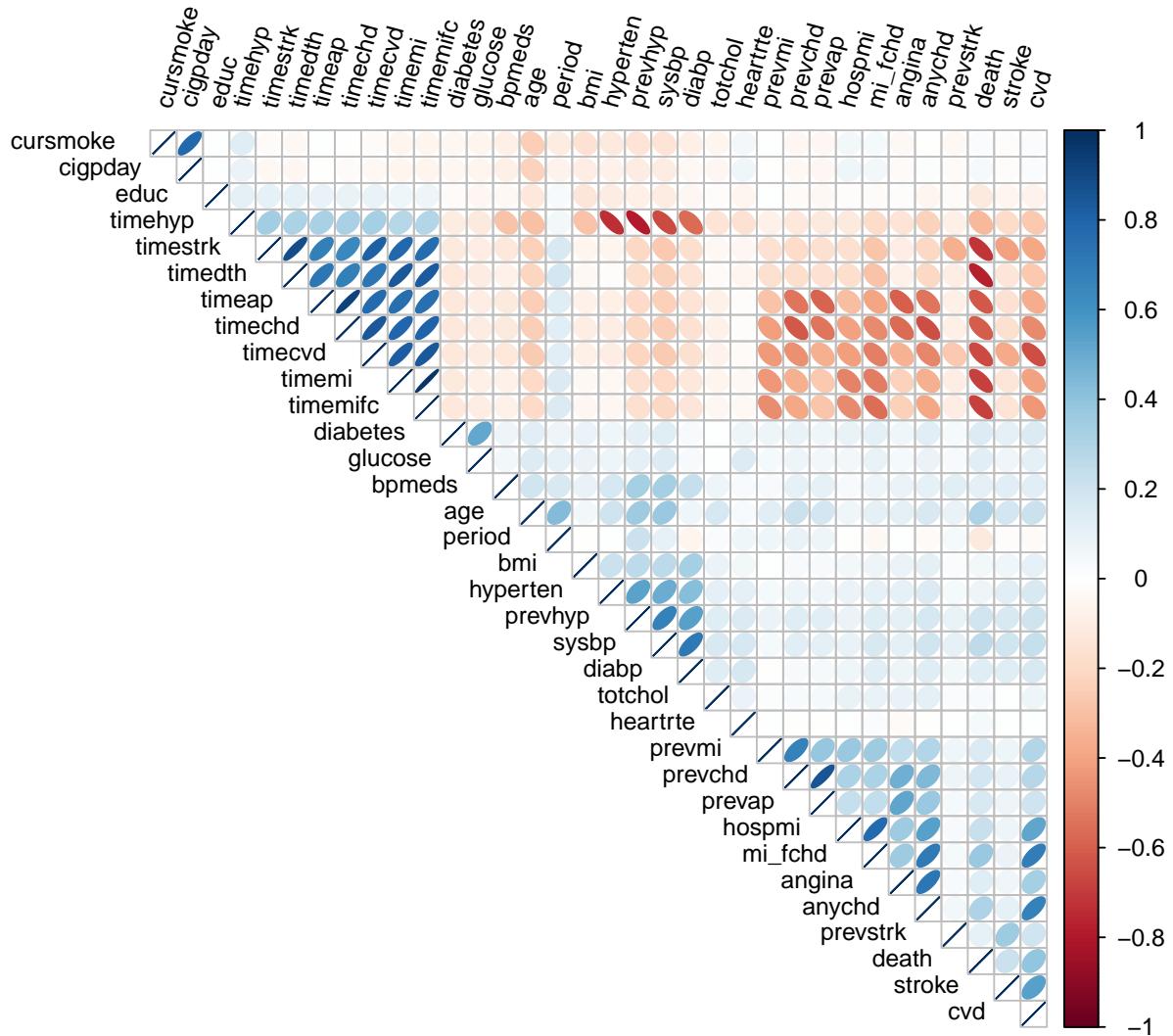
smoking ~ hyperten [highly sig]

```
c(chisq.test(table(dat$hyperten[dat$period == 1], dat$cursmoke[dat$period == 1]))$p.value,
  chisq.test(table(dat$hyperten, dat$cursmoke))$p.value)
```

```
## [1] 9.095863e-15 1.497426e-35
```

pairwise correlations

```
dat[,-c(1,2, 20, 38)] %>%
  cor(., use = "complete.obs") %>%
  corrplot::corrplot(., type = "upper", order = "hclust",
    tl.col = "black", tl.srt = 75, tl.offset = 1, tl.cex = .8, method = "ellipse")
```



demographics

education

```
dat %>%
  group_by(educ) %>%
  summarise(mean_resp = mean(cursmoke),
            lower = prop.test(sum(cursmoke), n())$conf.int[1],
            upper = prop.test(sum(cursmoke), n())$conf.int[2]) %>%
  ggplot(aes(x = educ, y = mean_resp)) +
  geom_errorbar(aes(ymin=lower, ymax=upper), width=.1) +
  geom_line() +
  theme_bw() +
  labs(title = 'Proportion of smokers by education (mean and 95% CI)', caption = "Figure 2")
```

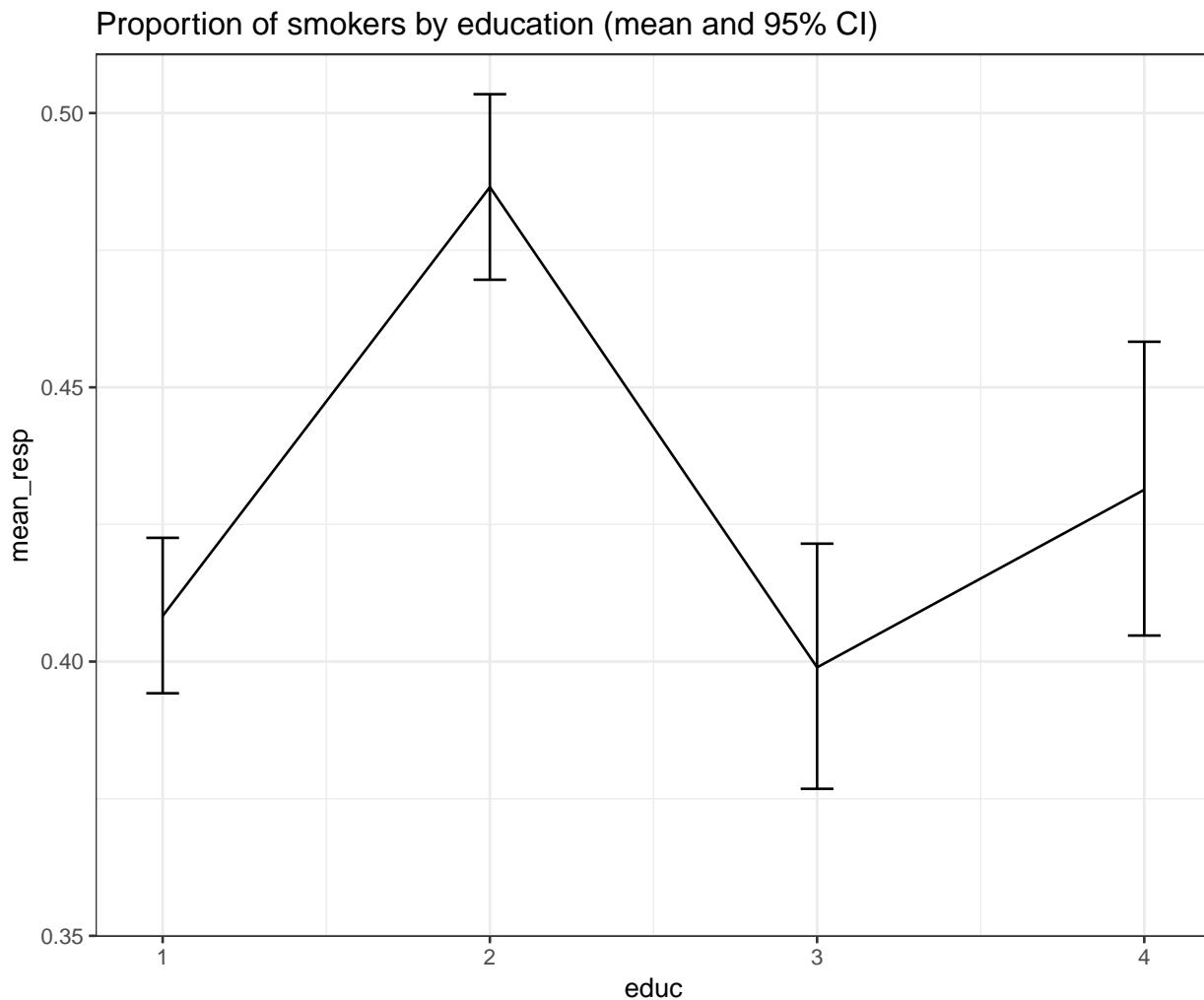


Figure 2

Model

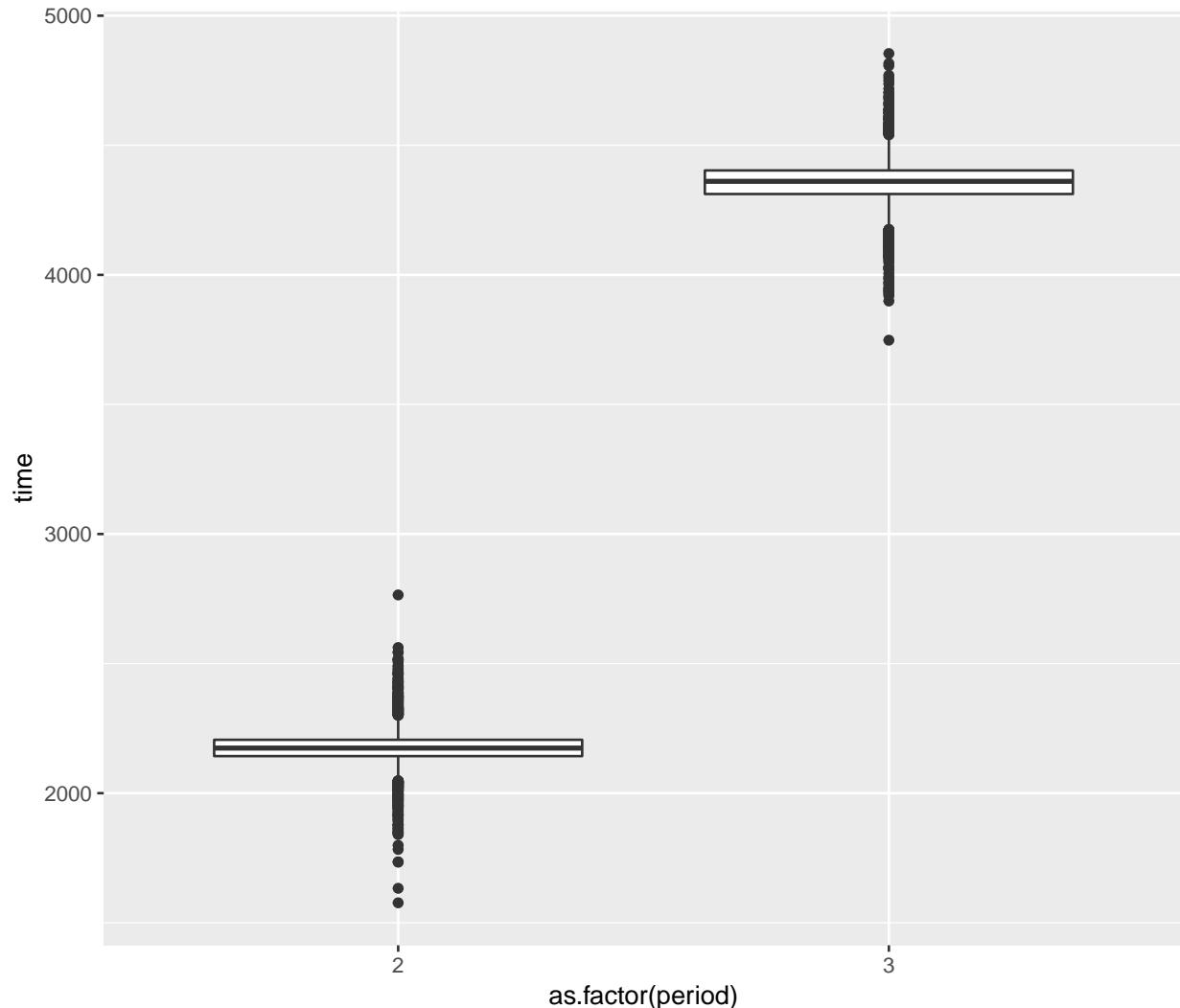
Smoking ~ age + sex + covariates

```
dat %>%
  na.omit() %>%
  group_by(randid) %>%
  mutate(num_visits = n()) %>%
  ungroup() %>%
  dplyr::select(randid, num_visits) %>%
  distinct() %>%
  summarise(mean(num_visits), sd(num_visits),
            max(num_visits), min(num_visits), median(num_visits)) %>%
  kable(caption = "Number of observations per subject stats") %>%
  kableExtra::kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Table 1: Number of observations per subject stats

mean(num_visits)	sd(num_visits)	max(num_visits)	min(num_visits)	median(num_visits)
2.209827	0.7296647	3	1	2

```
dat %>%
  filter(time > 0) %>%
  ggplot(aes(x = as.factor(period), y = time)) +
  geom_boxplot()
```



Motivation for conditional model:

- Incomplete data
- imbalanced number of observations per person: after missing data is removed the mean/median number of observations per subject is 2.2/2 with $sd = 0.7$. The minimum number of visits is 1 and the maximum is 3 (**Table above**).

- Unequally spaced f/up time: the spread of a distribution of number of days from the 1 st visit is within 200 days (**Boxplot above**).

demographics

```
m1 = lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + (1|randid), family = binomial, data = dat, nAGQ = 0)
m2 = lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + bmi + (1|randid), family = binomial, data = dat, nAGQ = 0)

summary(m1)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##   Family: binomial  ( logit )
##   Formula: cursmoke ~ age + sex + as.factor(educ) + (1 | randid)
##   Data: dat
##
```

```

##      AIC      BIC logLik deviance df.resid
##  10717.9 10769.3 -5352.0 10703.9     11325
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.3391 -0.2254 -0.1185  0.2316  2.3812
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   randid (Intercept) 21.26    4.611
## Number of obs: 11332, groups: randid, 4321
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      7.728999  0.393600 19.637 < 2e-16 ***
## age             -0.135732  0.006396 -21.222 < 2e-16 ***
## sex2            -1.236751  0.166942 -7.408 1.28e-13 ***
## as.factor(educ)2  0.025475  0.198878  0.128  0.8981
## as.factor(educ)3 -0.350353  0.240733 -1.455  0.1456
## as.factor(educ)4 -0.470844  0.271508 -1.734  0.0829 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   as.()2 as.()3
## age       -0.920
## sex2      -0.237  0.018
## as.fctr(d)2 -0.317  0.135 -0.062
## as.fctr(d)3 -0.226  0.083 -0.094  0.355
## as.fctr(d)4 -0.239  0.078  0.076  0.305  0.247
summary(m2)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula: cursmoke ~ age + sex + as.factor(educ) + bmi + (1 | randid)
## Data: dat
##
##      AIC      BIC logLik deviance df.resid
##  10477.3 10536.0 -5230.7 10461.3     11274
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.6155 -0.2210 -0.1139  0.2293  2.3928
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   randid (Intercept) 20.84    4.565
## Number of obs: 11282, groups: randid, 4307
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      11.968016  0.626679 19.098 < 2e-16 ***
## age             -0.133952  0.006434 -20.821 < 2e-16 ***

```

```

## sex2           -1.330784   0.167368  -7.951 1.85e-15 ***
## as.factor(educ)2 -0.111516   0.199642  -0.559  0.5764
## as.factor(educ)3 -0.559513   0.241297  -2.319  0.0204 *
## as.factor(educ)4 -0.661328   0.271375  -2.437  0.0148 *
## bmi            -0.162158   0.018212  -8.904 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   as.()2 as.()3 as.()4
## age      -0.578
## sex2     -0.212  0.019
## as.fctr(d)2 -0.260  0.134 -0.057
## as.fctr(d)3 -0.218  0.081 -0.088  0.362
## as.fctr(d)4 -0.214  0.076  0.080  0.312  0.255
## bmi      -0.776 -0.004  0.084  0.079  0.099  0.083
AIC(m1, m2)

##      df      AIC
## m1    7 10717.93
## m2    8 10477.34
# m2 is better

```

heart

```

###  

# adding heart variables: heartrte, prevstrk, totchol  

m3a = lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + bmi + heartrte + (1|randid), family = binomial)  

# m3b didnt converge  

m3b= lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + bmi + as.factor(prevstrk) + (1|randid), family = binomial)  

m3c =  

dat %>%  

select(randid, cursmoke, age, sex, educ, bmi, totchol) %>%  

na.omit() %>%  

lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + bmi + log(totchol) + (1|randid), family = binomial)  

AIC(m2, m3a, m3b, m3c)

##      df      AIC
## m2    8 10477.34
## m3a   9 10425.79
## m3b   9 10478.53
## m3c   9 10229.92
summary(m3a)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )

```

```

## Formula: cursmoke ~ age + sex + as.factor(educ) + bmi + heartrte + (1 |
##     randid)
## Data: dat
##
##      AIC      BIC logLik deviance df.resid
## 10425.8 10491.8 -5203.9 10407.8     11271
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.7943 -0.2223 -0.1126  0.2298  2.6208
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.66    4.545
## Number of obs: 11280, groups: randid, 4307
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 10.754161  0.678719 15.845 < 2e-16 ***
## age         -0.135968  0.006455 -21.065 < 2e-16 ***
## sex2        -1.382799  0.167445 -8.258 < 2e-16 ***
## as.factor(educ)2 -0.129696  0.199196 -0.651  0.5150
## as.factor(educ)3 -0.551866  0.240657 -2.293  0.0218 *
## as.factor(educ)4 -0.626044  0.270845 -2.311  0.0208 *
## bmi          -0.166939  0.018235 -9.155 < 2e-16 ***
## heartrte     0.019162  0.004251  4.507 6.57e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) age     sex2   as.()2 as.()3 as.()4 bmi
## age         -0.499
## sex2        -0.168  0.025
## as.fctr(d)2 -0.230  0.136 -0.055
## as.fctr(d)3 -0.201  0.080 -0.088  0.363
## as.fctr(d)4 -0.207  0.074  0.078  0.311  0.256
## bmi          -0.691  0.001  0.089  0.079  0.097  0.081
## heartrte    -0.386 -0.085 -0.073 -0.025  0.002  0.025 -0.061
summary(m3b)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + as.factor(educ) + bmi + as.factor(prevstrk) +
##   (1 | randid)
## Data: dat
##
##      AIC      BIC logLik deviance df.resid
## 10478.5 10544.5 -5230.3 10460.5     11273
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.6132 -0.2211 -0.1139  0.2289  2.6684

```

```

## 
## Random effects:
## Groups Name           Variance Std.Dev.
## randid (Intercept) 20.85     4.566
## Number of obs: 11282, groups: randid, 4307
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                11.96044   0.62690 19.079 < 2e-16 ***
## age                      -0.13356   0.00646 -20.677 < 2e-16 ***
## sex2                     -1.33123   0.16741 -7.952 1.84e-15 ***
## as.factor(educ)2          -0.11170   0.19968 -0.559   0.5759
## as.factor(educ)3          -0.56077   0.24136 -2.323   0.0202 *
## as.factor(educ)4          -0.66374   0.27144 -2.445   0.0145 *
## bmi                      -0.16248   0.01822 -8.917 < 2e-16 ***
## as.factor(prevstrk)1      -0.36229   0.51834 -0.699   0.4846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) age    sex2   as.()2 as.()3 as.()4 bmi
## age       -0.577
## sex2      -0.212  0.018
## as.fctr(d)2 -0.260  0.134 -0.057
## as.fctr(d)3 -0.217  0.080 -0.088  0.362
## as.fctr(d)4 -0.214  0.074  0.080  0.312  0.255
## bmi        -0.775 -0.006  0.084  0.079  0.099  0.084
## as.fctr(p)1  0.019 -0.086  0.003  0.001  0.007  0.012  0.024

```

```
summary(m3c)
```

```

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula: cursmoke ~ age + sex + as.factor(educ) + bmi + log(totchol) +
##           (1 | randid)
## Data: .
##
##      AIC      BIC      logLik deviance df.resid
## 10229.9 10295.6  -5106.0  10211.9      10887
##
## Scaled residuals:
##      Min      1Q      Median      3Q      Max
## -2.5138 -0.2285 -0.1137  0.2309  2.5367
##
## Random effects:
## Groups Name           Variance Std.Dev.
## randid (Intercept) 20.48     4.526
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                6.074642   1.800070  3.375 0.000739 ***
## age                      -0.134748   0.006555 -20.558 < 2e-16 ***
## sex2                     -1.405960   0.167925 -8.373 < 2e-16 ***

```

```

## as.factor(educ)2 -0.121810  0.199207 -0.611 0.540885
## as.factor(educ)3 -0.575254  0.240626 -2.391 0.016818 *
## as.factor(educ)4 -0.655196  0.270075 -2.426 0.015267 *
## bmi             -0.170700  0.018649 -9.153 < 2e-16 ***
## log(totchol)    1.131222  0.329470  3.433 0.000596 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age   sex2  as.()2 as.()3 as.()4 bmi
## age      -0.106
## sex2     0.030  0.032
## as.fctr(d)2 -0.087  0.135 -0.055
## as.fctr(d)3 -0.056  0.083 -0.085  0.363
## as.fctr(d)4 -0.062  0.076  0.081  0.313  0.256
## bmi       -0.137  0.006  0.102  0.079  0.100  0.085
## log(totchl) -0.936 -0.101 -0.112 -0.005 -0.022 -0.014 -0.142
# totchol has the best AIC

```

blood pressure - nothing is significant

Check blood pressure predictors: systolic, diastolic bp, bp medications, hypertantion, prevhyp

```

# dichotomize education into high school and college
dat$educ_binary <- NA
dat$educ_binary[dat$educ %in% c(1,2)] <- "high_school"
dat$educ_binary[dat$educ %in% c(3,4)] <- "college"

table(dat$educ)

##
##      1     2     3     4
## 4690 3410 1885 1347
table(dat$educ_binary)

##
##      college high_school
##      3232      8100

dat = dat %>% mutate(educ_binary = factor(educ_binary))

m3c =
  dat %>%
  select(randid, cursmoke, age, sex, educ, bmi, totchol) %>%
  na.omit() %>%
  lme4::glmer(cursmoke ~ age + sex + as.factor(educ) + bmi + log(totchol) + (1|randid), family = binomial)

m3c1 =
  dat %>%
  select(randid, cursmoke, age, sex, educ_binary, bmi, totchol) %>%
  na.omit() %>%
  lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + (1|randid), family = binomial, c

```

```

summary(m3c1)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula: cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + (1 |
##         randid)
## Data: .
##
##      AIC      BIC  logLik deviance df.resid
## 10224.6 10275.7 -5105.3 10210.6    10889
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -2.5109 -0.2281 -0.1133  0.2309  2.5380
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.57    4.536
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.424064  1.797310  3.018 0.002546 **
## age        -0.134412  0.006502 -20.673 < 2e-16 ***
## sex2       -1.407104  0.166504 -8.451 < 2e-16 ***
## educ_binaryhigh_school 0.556860  0.182191  3.056 0.002240 **
## bmi        -0.169991  0.018624 -9.128 < 2e-16 ***
## log(totchol) 1.132360  0.329910  3.432 0.000598 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age   sex2   edc_b_ bmi
## age      -0.091
## sex2      0.028  0.039
## edc_bnryhg_-0.061 -0.047 -0.012
## bmi      -0.121 -0.005  0.108 -0.093
## log(totchl) -0.942 -0.101 -0.114  0.023 -0.142

```

```
AIC(m3c, m3c1)
```

```

##      df      AIC
## m3c    9 10229.92
## m3c1   7 10224.59

```

```
m4a= lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + sysbp + (1|randid), family =
```

```
m4b= lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + diabp + (1|randid), family =
```

```
m4c= lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + hyperten + (1|randid), family =
```

```

m4d= lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevhyp + (1|randid), family = "binomial")
m4e= lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + bpmeds + (1|randid), family = "binomial")

AIC(m4a, m4b, m4c, m4d, m4e)

##      df      AIC
## m4a   8 10226.000
## m4b   8 10226.119
## m4c   8 10216.501
## m4d   8 10225.609
## m4e   8  9893.857
summary(m4a)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##   Family: binomial  ( logit )
##   Formula:
##   cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + sysbp +
##           (1 | randid)
##   Data: dat
##
##      AIC      BIC  logLik deviance df.resid
##  10226.0 10284.4 -5105.0 10210.0     10888
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.5086 -0.2282 -0.1134  0.2307  2.6016
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   randid (Intercept) 20.58     4.537
##   Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.439936  1.797871  3.026  0.00248 ***
## age        -0.133149  0.006939 -19.190 < 2e-16 ***
## sex2       -1.404386  0.166621 -8.429 < 2e-16 ***
## educ_binaryhigh_school 0.559555  0.182295  3.070  0.00214 **
## bmi        -0.167773  0.019101 -8.784 < 2e-16 ***
## log(totchol) 1.144395  0.330761  3.460  0.00054 ***
## sysbp      -0.001549  0.002941 -0.527  0.59838
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age        -0.079
## sex2       0.029  0.048
## edc_bnryhg_-0.060 -0.035 -0.011
## bmi        -0.115  0.072  0.112 -0.084
## log(totchl)-0.938 -0.071 -0.112  0.025 -0.123

```

```

## sysbp      -0.016 -0.349 -0.032 -0.028 -0.221 -0.069
summary(m4b)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial  ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + diabp +
##           (1 | randid)
## Data: dat
##
##          AIC      BIC  logLik deviance df.resid
## 10226.1 10284.5 -5105.1 10210.1     10888
## 
## Scaled residuals:
##    Min     1Q  Median     3Q    Max
## -2.5153 -0.2280 -0.1135  0.2306  2.5418
## 
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.56     4.535
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                5.434094  1.797286  3.024 0.002499 ***
## age                     -0.134369  0.006501 -20.670 < 2e-16 ***
## sex2                    -1.408742  0.166573 -8.457 < 2e-16 ***
## educ_binaryhigh_school  0.557039  0.182158  3.058 0.002228 **
## bmi                      -0.168395  0.019308 -8.722 < 2e-16 ***
## log(totchol)              1.146884  0.333287  3.441 0.000579 ***
## diabp                   -0.001585  0.005137 -0.309 0.757667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age         -0.091
## sex2        0.028  0.039
## edc_bnryhg_-0.060 -0.047 -0.012
## bmi         -0.112 -0.003  0.095 -0.088
## log(totchl)-0.929 -0.099 -0.118  0.023 -0.098
## diabp      -0.019 -0.008  0.035 -0.005 -0.264 -0.143
summary(m4c)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial  ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + hyperten +
##           (1 | randid)
## Data: dat
##

```

```

##      AIC      BIC logLik deviance df.resid
##  10216.5 10274.9 -5100.3 10200.5     10888
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.4940 -0.2280 -0.1137  0.2308  2.5329
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   randid (Intercept) 20.5      4.527
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)               5.331142  1.796471  2.968 0.003002 **
## age                   -0.133249  0.006537 -20.385 < 2e-16 ***
## sex2                  -1.402311  0.166332 -8.431 < 2e-16 ***
## educ_binaryhigh_school 0.568665  0.182228  3.121 0.001805 **
## bmi                   -0.165490  0.018894 -8.759 < 2e-16 ***
## log(totchol)            1.147616  0.329820  3.480 0.000502 ***
## hyperten                -0.246507  0.189591 -1.300 0.193533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age        -0.094
## sex2       0.028  0.041
## edc_bnryhg_-0.062 -0.041 -0.011
## bmi        -0.126  0.014  0.109 -0.082
## log(totchl) -0.942 -0.096 -0.113  0.025 -0.133
## hyperten    0.037 -0.115 -0.016 -0.053 -0.177 -0.039
summary(m4d)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevhyp +
## (1 | randid)
## Data: dat
##
##      AIC      BIC logLik deviance df.resid
##  10225.6 10284.0 -5104.8 10209.6     10888
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.4758 -0.2283 -0.1132  0.2318  2.5841
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   randid (Intercept) 20.52      4.53
## Number of obs: 10896, groups: randid, 4293
##

```

```

## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)            5.168962   1.811973  2.853 0.004335 **
## age                  -0.131314   0.007072 -18.568 < 2e-16 ***
## sex2                 -1.408191   0.166305 -8.468 < 2e-16 ***
## educ_binaryhigh_school 0.558072   0.181965  3.067 0.002163 **
## bmi                  -0.166579   0.018866 -8.829 < 2e-16 ***
## log(totchol)          1.143262   0.329982  3.465 0.000531 ***
## prevhyp              -0.135577   0.128181 -1.058 0.290191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Correlation of Fixed Effects:
##             (Intr) age    sex2  edc_b_ bmi   lg(tt)
## age         -0.134
## sex2        0.030  0.032
## edc_bnryhg_-0.061 -0.040 -0.012
## bmi         -0.140  0.060  0.105 -0.090
## log(totchl) -0.938 -0.080 -0.114  0.023 -0.135
## prevhyp     0.130 -0.394  0.012 -0.009 -0.165 -0.034

```

```
summary(m4e)
```

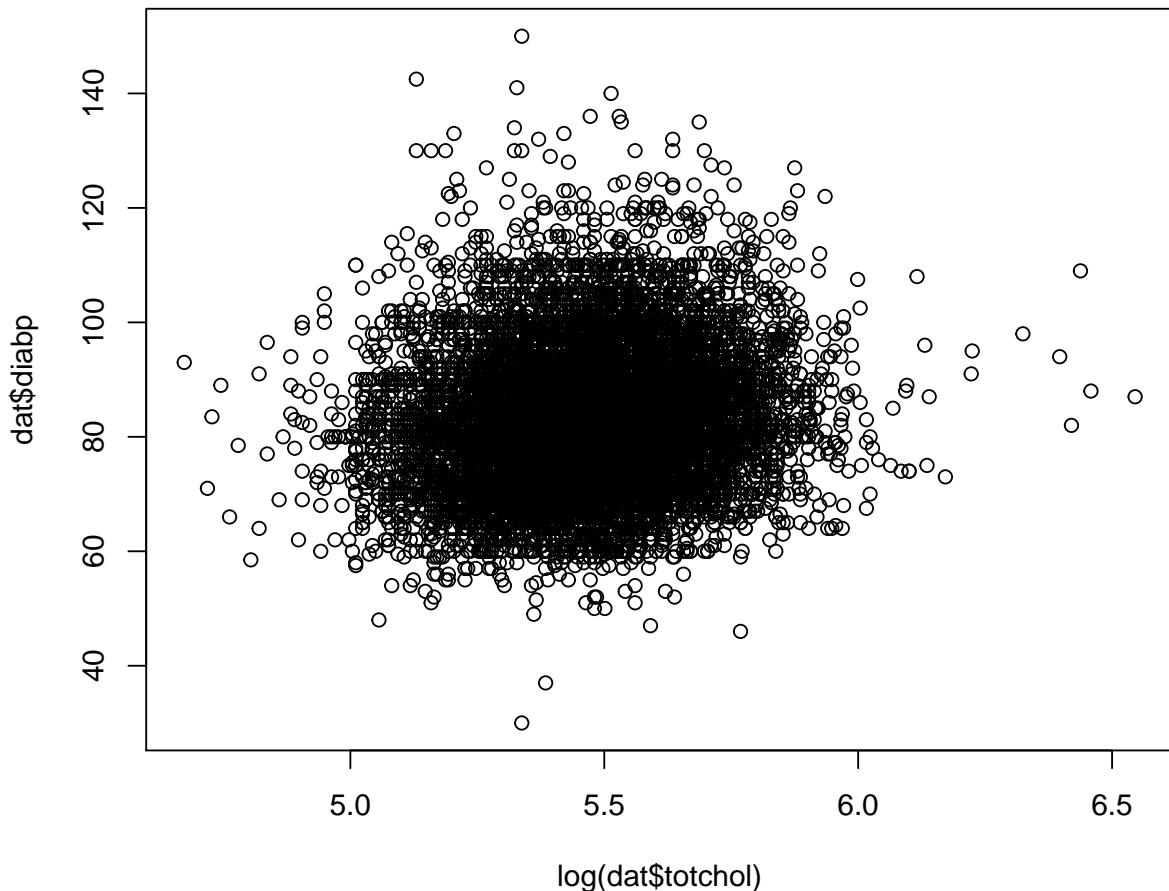
```

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial  ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + bpmeds +
##           (1 | randid)
## Data: dat
##
##      AIC      BIC      logLik deviance df.resid
## 9893.9  9951.8  -4938.9   9877.9     10313
##
## Scaled residuals:
##      Min      1Q      Median      3Q      Max
## -2.4666 -0.2353 -0.1193  0.2398  2.3496
##
## Random effects:
## Groups Name      Variance Std.Dev.
## randid (Intercept) 19.84     4.454
## Number of obs: 10321, groups: randid, 4280
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)            5.512823   1.819273  3.030 0.00244 **
## age                  -0.131096   0.006802 -19.272 < 2e-16 ***
## sex2                 -1.405928   0.164909 -8.525 < 2e-16 ***
## educ_binaryhigh_school 0.538936   0.180106  2.992 0.00277 **
## bmi                  -0.163186   0.018671 -8.740 < 2e-16 ***
## log(totchol)          1.052086   0.335038  3.140 0.00169 **
## bpmeds               0.050145   0.195748  0.256 0.79782
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Correlation of Fixed Effects:
##          (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age      -0.077
## sex2      0.026  0.054
## edc_bnryhg_-0.059 -0.046 -0.011
## bmi       -0.122  0.006  0.111 -0.093
## log(totchl) -0.941 -0.123 -0.113  0.023 -0.141
## bpmeds     0.044 -0.224 -0.051  0.002 -0.031  0.003
plot(log(dat$totchol), dat$diabp)

```

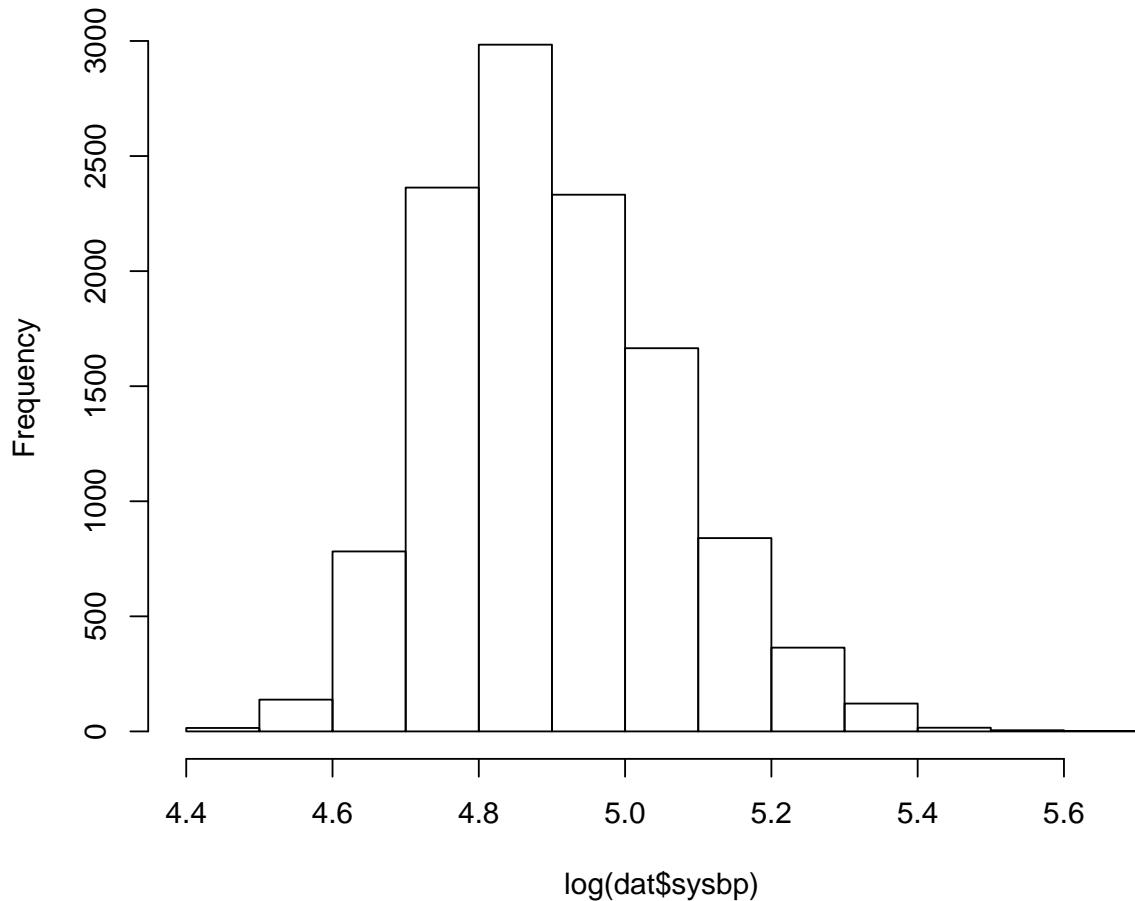


```

#hist(dat$sysbp)
hist(log(dat$sysbp))

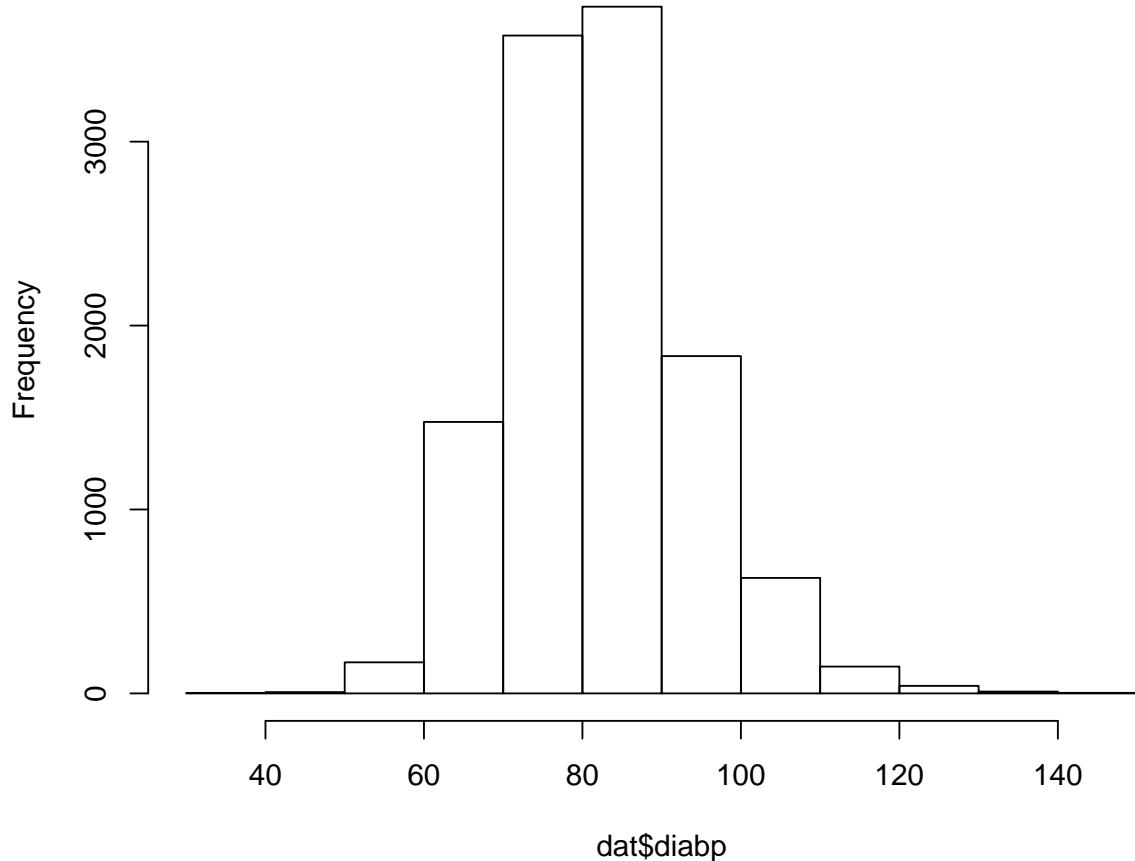
```

Histogram of $\log(\text{dat}\$sysbp)$



```
hist(dat$diabp)
```

Histogram of dat\$diabp



```
#hist(log(dat$diabp))
```

diabetes - nothing

Check diabetes and glucose

```
#start model
m3c1 =
  dat %>%
    select(randid, cursmoke, age, sex, educ_binary, bmi, totchol) %>%
    na.omit() %>%
    lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + (1|randid), family = binomial,
                REML = FALSE)
m5a = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + diabetes + (1|randid), family = binomial,
                   REML = FALSE)
m5b = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + log(glucose) + (1|randid), family = binomial,
                   REML = FALSE)

AIC(m5a, m5b)
```

```

##      df      AIC
## m5a   8 10226.532
## m5b   8  9500.351
summary(m5a)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + diabetes +
##          (1 | randid)
## Data: dat
##
##      AIC      BIC  logLik deviance df.resid
## 10226.5 10284.9 -5105.3 10210.5     10888
##
## Scaled residuals:
##    Min     1Q  Median     3Q    Max
## -2.5081 -0.2282 -0.1132  0.2309  2.5403
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.57    4.536
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.430852  1.797550  3.021 0.002517 **
## age         -0.134113  0.006604 -20.307 < 2e-16 ***
## sex2        -1.407409  0.166503 -8.453 < 2e-16 ***
## educ_binaryhigh_school 0.556784  0.182183  3.056 0.002242 **
## bmi         -0.169883  0.018629 -9.119 < 2e-16 ***
## log(totchol) 1.128261  0.330301  3.416 0.000636 ***
## diabetes    -0.069308  0.272311 -0.255 0.799097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age         -0.087
## sex2        0.028  0.037
## edc_bnryhg_-0.061 -0.047 -0.012
## bmi         -0.121 -0.001  0.108 -0.093
## log(totchl)-0.941 -0.108 -0.113  0.023 -0.143
## diabetes   -0.015 -0.176  0.008  0.001 -0.022  0.048
summary(m5b)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + log(glucose) +
##          (1 | randid)

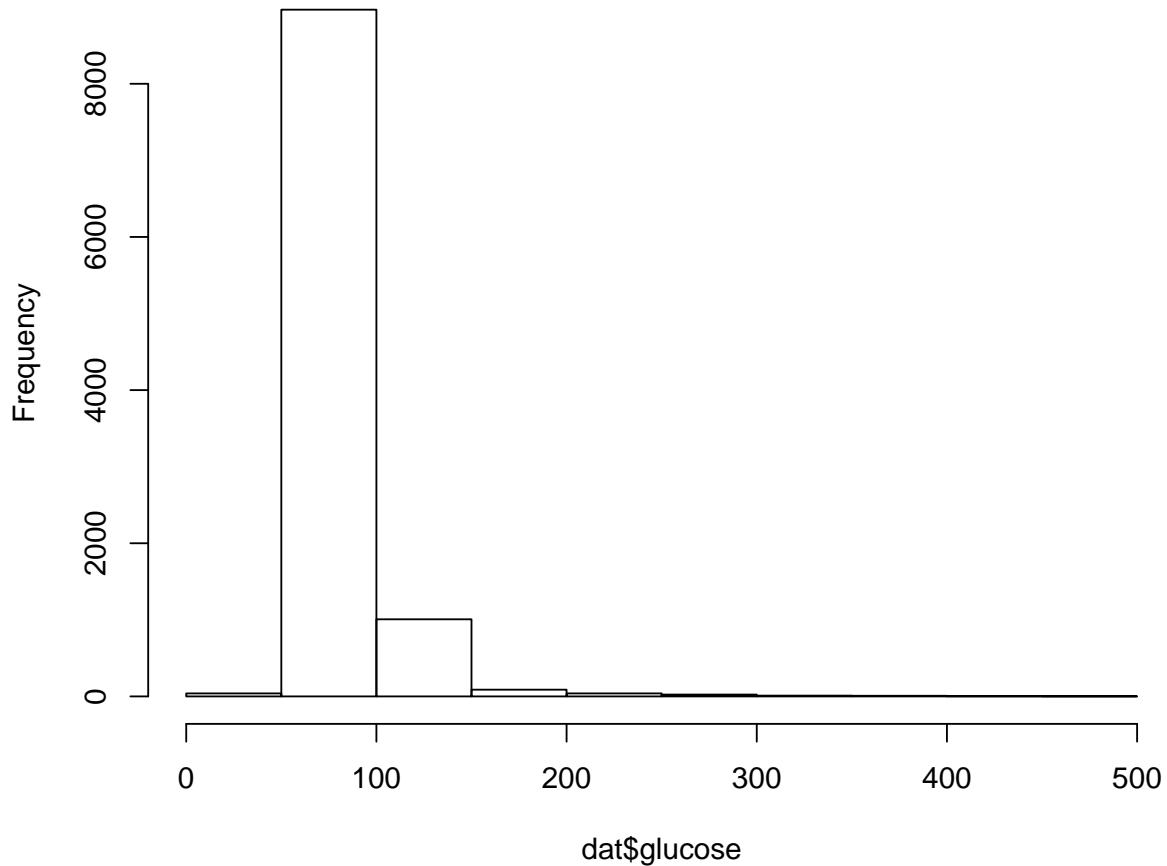
```

```

##      Data: dat
##
##      AIC      BIC  logLik deviance df.resid
##  9500.4   9557.8 -4742.2   9484.4     9753
##
## Scaled residuals:
##      Min    1Q Median    3Q   Max
## -2.4703 -0.2406 -0.1254  0.2451  2.7037
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 19.35    4.399
## Number of obs: 9761, groups: randid, 4236
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                7.291710  2.166857  3.365 0.000765 ***
## age                      -0.124136  0.006854 -18.110 < 2e-16 ***
## sex2                     -1.378019  0.164360 -8.384 < 2e-16 ***
## educ_binaryhigh_school   0.516307  0.179991  2.869 0.004124 **
## bmi                      -0.160842  0.018792 -8.559 < 2e-16 ***
## log(totchol)              0.960902  0.340141  2.825 0.004728 **
## log(glucose)             -0.385391  0.247717 -1.556 0.119762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age       0.035
## sex2      0.027  0.036
## edc_bnryhg_-0.043 -0.043 -0.011
## bmi       -0.074  0.010  0.105 -0.092
## log(totchl) -0.842 -0.126 -0.115  0.022 -0.139
## log(glucos) -0.524 -0.197  0.004 -0.013 -0.061  0.080
hist(dat$glucose)

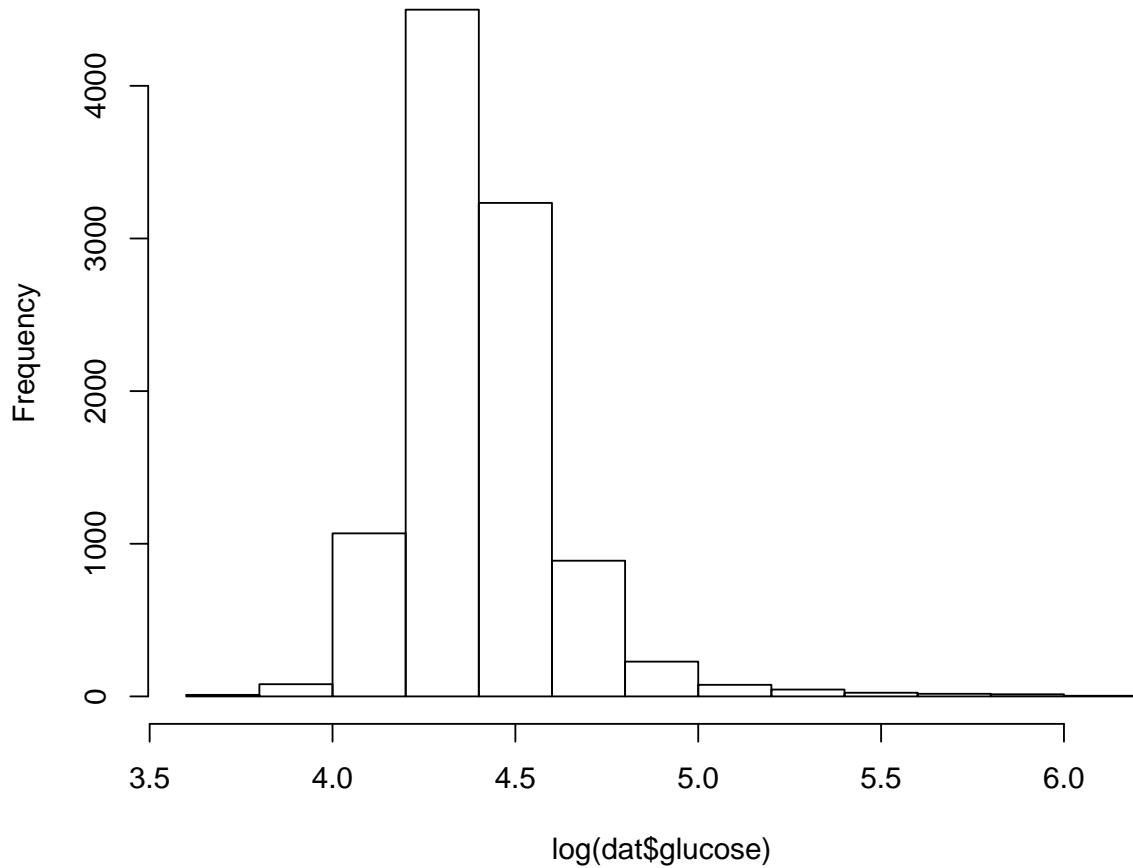
```

Histogram of dat\$glucose



```
hist(log(dat$glucose))
```

Histogram of log(dat\$glucose)



Prevalence

```
m3c1 =  
  dat %>%  
    select(randid, cursmoke, age, sex, educ_binary, bmi, totchol) %>%  
    na.omit() %>%  
  lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + (1|randid), family = binomial, c  
  
m6a = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + mi_fchd + (1|randid), family = binomial, c  
  
m6b = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevchd + (1|randid), family = binomial, c  
  
m6c = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevap + (1|randid), family = binomial, c  
  
m6d = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + death + (1|randid), family = binomial, c  
  
m6e = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + angina + (1|randid), family = binomial, c
```

```

m6f = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + hospmi + (1|randid), family =
m6g = lme4::glmer(cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + cvd + (1|randid), family =
AIC(m6a, m6b, m6c, m6d, m6e, m6f, m6g)

##      df      AIC
## m6a   8 10212.04
## m6b   8 10220.44
## m6c   8 10219.51
## m6d   8 10177.80
## m6e   8 10226.58
## m6f   8 10217.41
## m6g   8 10202.94

summary(m6a) # mi_fchd

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial  ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + mi_fchd +
##       (1 | randid)
## Data: dat
##
##      AIC      BIC      logLik deviance df.resid
## 10212.0 10270.4 -5098.0  10196.0     10888
##
## Scaled residuals:
##      Min      1Q      Median      3Q      Max
## -2.5279 -0.2285 -0.1143  0.2317  2.5429
##
## Random effects:
## Groups Name      Variance Std.Dev.
## randid (Intercept) 20.42    4.519
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.799923  1.800387  3.221  0.00128 **
## age        -0.135581  0.006519 -20.796 < 2e-16 ***
## sex2        -1.303790  0.169643 -7.685 1.52e-14 ***
## educ_binaryhigh_school 0.556624  0.181624  3.065  0.00218 **
## bmi         -0.171769  0.018594 -9.238 < 2e-16 ***
## log(totchol) 1.054103  0.330480  3.190  0.00142 **
## mi_fchd     0.648221  0.228313  2.839  0.00452 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age     sex2    edc_b_ bmi     lg(tt)
## age        -0.098
## sex2        0.044  0.021
## edc_bnryhg_-0.060 -0.048 -0.011

```

```

## bmi          -0.125 -0.002  0.097 -0.093
## log(totchl) -0.942 -0.093 -0.128  0.022 -0.137
## mi_fchd      0.077 -0.085  0.206  0.003 -0.042 -0.080

summary(m6b) # prevchd

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##   Family: binomial ( logit )
## Formula:
##   cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevchd +
##             (1 | randid)
## Data: dat
##
##       AIC     BIC   logLik deviance df.resid
##   10220.4 10278.8 -5102.2 10204.4    10888
##
## Scaled residuals:
##   Min     1Q   Median     3Q    Max
## -2.4845 -0.2278 -0.1133  0.2293  2.5016
##
## Random effects:
##   Groups Name        Variance Std.Dev.
##   randid (Intercept) 20.75    4.555
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.446805  1.801507  3.023 0.002499 **
## age            -0.131261  0.006691 -19.619 < 2e-16 ***
## sex2           -1.441790  0.167816 -8.592 < 2e-16 ***
## educ_binaryhigh_school 0.556313  0.182906  3.042 0.002354 **
## bmi            -0.171430  0.018714 -9.161 < 2e-16 ***
## log(totchol)    1.113985  0.330785  3.368 0.000758 ***
## prevchd        -0.508868  0.226456 -2.247 0.024634 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age         -0.088
## sex2        0.028  0.018
## edc_bnryhg_-0.061 -0.047 -0.011
## bmi         -0.122 -0.012  0.111 -0.092
## log(totchl)-0.941 -0.104 -0.111  0.023 -0.142
## prevchd    -0.003 -0.224  0.089  0.003  0.029  0.026

summary(m6c) # prevap

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##   Family: binomial ( logit )
## Formula:
##   cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + prevap +
##             (1 | randid)

```

```

##      Data: dat
##
##      AIC      BIC  logLik deviance df.resid
##  10219.5 10277.9 -5101.8 10203.5     10888
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.4899 -0.2277 -0.1129  0.2293  2.5528
##
## Random effects:
## Groups Name      Variance Std.Dev.
## randid (Intercept) 20.75    4.555
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                5.384395   1.802154   2.988 0.002810 **
## age                     -0.131927   0.006642 -19.863 < 2e-16 ***
## sex2                    -1.434434   0.167613  -8.558 < 2e-16 ***
## educ_binaryhigh_school  0.557229   0.182929   3.046 0.002318 **
## bmi                      -0.171460   0.018704  -9.167 < 2e-16 ***
## log(totchol)              1.130002   0.330801   3.416 0.000636 ***
## prevap                  -0.574383   0.266989  -2.151 0.031450 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age        -0.092
## sex2       0.029  0.025
## edc_bnryhg_-0.060 -0.047 -0.012
## bmi        -0.121 -0.011  0.111 -0.093
## log(totchl) -0.941 -0.101 -0.113  0.023 -0.143
## prevap      0.012 -0.190  0.071  0.001  0.031  0.006
summary(m6d) # death

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + death +
## (1 | randid)
## Data: dat
##
##      AIC      BIC  logLik deviance df.resid
##  10177.8 10236.2 -5080.9 10161.8     10888
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.5691 -0.2312 -0.1113  0.2318  2.6425
##
## Random effects:
## Groups Name      Variance Std.Dev.
## randid (Intercept) 19.95    4.467

```

```

## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)             5.685667  1.788127  3.180  0.00147 **
## age                  -0.142030  0.006689 -21.235 < 2e-16 ***
## sex2                 -1.247278  0.166200 -7.505 6.16e-14 ***
## educ_binaryhigh_school 0.483921  0.180173  2.686  0.00723 **
## bmi                  -0.170943  0.018441 -9.270 < 2e-16 ***
## log(totchol)          1.095348  0.328015  3.339  0.00084 ***
## death                 1.027647  0.179417  5.728 1.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age         -0.096
## sex2        0.034  0.001
## edc_bnryhg_-0.061 -0.032 -0.021
## bmi         -0.123  0.002  0.102 -0.092
## log(totchl) -0.942 -0.096 -0.116  0.024 -0.140
## death       0.033 -0.244  0.151 -0.065 -0.026 -0.013
summary(m6f) # hospmi

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + hospmi +
##   (1 | randid)
## Data: dat
##
##      AIC      BIC      logLik deviance df.resid
## 10217.4 10275.8  -5100.7  10201.4     10888
##
## Scaled residuals:
##      Min      1Q      Median      3Q      Max
## -2.5195 -0.2280 -0.1141  0.2311  2.5364
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.49     4.526
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)             5.676586  1.800065  3.154  0.00161 **
## age                  -0.134925  0.006507 -20.734 < 2e-16 ***
## sex2                 -1.339143  0.169022 -7.923 2.32e-15 ***
## educ_binaryhigh_school 0.558264  0.181879  3.069  0.00214 **
## bmi                  -0.170775  0.018604 -9.180 < 2e-16 ***
## log(totchol)          1.076842  0.330533  3.258  0.00112 **
## hospmi                0.591147  0.274581  2.153  0.03133 *
## ---

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age        -0.095
## sex2       0.040  0.029
## edc_bnryhg_-0.060 -0.047 -0.011
## bmi        -0.123 -0.004  0.101 -0.093
## log(totchl) -0.942 -0.096 -0.126  0.022 -0.139
## hospmi      0.068 -0.053  0.181  0.006 -0.025 -0.076
summary(m6g) # cvd

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula: cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + cvd +
##           (1 | randid)
## Data: dat
##
##      AIC      BIC      logLik deviance df.resid
## 10202.9 10261.3  -5093.5  10186.9     10888
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.6235 -0.2284 -0.1137  0.2324  2.5777
##
## Random effects:
## Groups Name      Variance Std.Dev.
## randid (Intercept) 20.25     4.5
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.845047  1.795846  3.255  0.00113 **
## age            -0.137605  0.006573 -20.935 < 2e-16 ***
## sex2           -1.282467  0.168032 -7.632 2.31e-14 ***
## educ_binaryhigh_school 0.528528  0.181158  2.918  0.00353 **
## bmi            -0.174270  0.018576 -9.381 < 2e-16 ***
## log(totchol)    1.062963  0.329273  3.228  0.00125 **
## cvd            0.755695  0.192811  3.919 8.88e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   edc_b_ bmi    lg(tt)
## age        -0.100
## sex2       0.040  0.010
## edc_bnryhg_-0.062 -0.042 -0.018
## bmi        -0.126  0.006  0.093 -0.091
## log(totchl) -0.942 -0.093 -0.121  0.024 -0.137
## cvd        0.065 -0.157  0.177 -0.036 -0.071 -0.049

```

```

summary(m6e) # angina(no)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ age + sex + educ_binary + bmi + log(totchol) + angina +
## (1 | randid)
## Data: dat
##
##      AIC      BIC  logLik deviance df.resid
## 10226.6 10284.9 -5105.3 10210.6     10888
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -2.5103 -0.2282 -0.1134  0.2311  2.5374
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 20.55     4.533
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                5.421864   1.800881   3.011 0.002607 ***
## age                     -0.134356   0.006516 -20.620 < 2e-16 ***
## sex2                    -1.406834   0.167076  -8.420 < 2e-16 ***
## educ_binaryhigh_school  0.556688   0.182113   3.057 0.002237 **
## bmi                      -0.169944   0.018642  -9.116 < 2e-16 ***
## log(totchol)              1.132039   0.330514   3.425 0.000615 ***
## angina                  -0.001081   0.224322  -0.005 0.996156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) age    sex2   edc_b_ bmi   lg(tt)
## age       -0.095
## sex2       0.034  0.033
## edc_bnryhg_-0.059 -0.048 -0.010
## bmi        -0.125 -0.001  0.103 -0.093
## log(totchl) -0.942 -0.096 -0.119  0.022 -0.138
## angina      0.068 -0.070  0.089  0.014 -0.053 -0.066

m6h = lme4::glmer(cursmoke ~ cage + sex + educ_binary + bmi + log(totchol) + prevchd + death + cvd + (1 | randid))

summary(m6h)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ cage + sex + educ_binary + bmi + log(totchol) + prevchd +
## death + cvd + (1 | randid)
## Data: dat

```

```

##
##      AIC      BIC logLik deviance df.resid
##  10162.8 10235.8 -5071.4 10142.8     10886
##
## Scaled residuals:
##      Min      1Q Median      3Q      Max
## -2.7465 -0.2311 -0.1101  0.2313  2.6835
##
## Random effects:
## Groups Name        Variance Std.Dev.
## randid (Intercept) 19.97    4.468
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0.842114  1.791766 -0.470  0.63836
## cage                   -0.138907  0.006838 -20.314 < 2e-16 ***
## sex2                  -1.227966  0.168032 -7.308 2.71e-13 ***
## educ_binaryhigh_school 0.471376  0.180382  2.613  0.00897 **
## bmi                    -0.175659  0.018546 -9.471 < 2e-16 ***
## log(totchol)            1.027651  0.328788  3.126  0.00177 **
## prevchd                -0.710396  0.232885 -3.050  0.00229 **
## death                  0.905432  0.191588  4.726 2.29e-06 ***
## cvd                     0.522970  0.207219  2.524  0.01161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) cage   sex2   edc_b_ bmi    lg(tt) prvchd death
## cage       0.091
## sex2      -0.037 -0.019
## edc_bnryhg_-0.068 -0.033 -0.023
## bmi        -0.128 -0.001  0.095 -0.091
## log(totchl) -0.960 -0.096 -0.119  0.024 -0.135
## prevchd    -0.047 -0.182  0.056  0.012  0.043  0.032
## death      -0.025 -0.194  0.092 -0.057 -0.002  0.004 -0.028
## cvd         0.050 -0.043  0.121 -0.016 -0.073 -0.051 -0.163 -0.339

```

Final model:

```
m6h = lme4::glmer(cursmoke ~ cage + sex + educ_binary + bmi + log(totchol) + prevchd + death + cvd + (1 | randid))

summary(m6h)
```

```

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cursmoke ~ cage + sex + educ_binary + bmi + log(totchol) + prevchd +
##   death + cvd + (1 | randid)
## Data: dat
##
##      AIC      BIC logLik deviance df.resid
##  10162.8 10235.8 -5071.4 10142.8     10886
##
```

```

## Scaled residuals:
##      Min     1Q Median     3Q    Max
## -2.7465 -0.2311 -0.1101  0.2313  2.6835
##
## Random effects:
##   Groups Name        Variance Std.Dev.
##   randid (Intercept) 19.97    4.468
## Number of obs: 10896, groups: randid, 4293
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.842114  1.791766 -0.470  0.63836
## cage                  -0.138907  0.006838 -20.314 < 2e-16 ***
## sex2                 -1.227966  0.168032  -7.308 2.71e-13 ***
## educ_binaryhigh_school 0.471376  0.180382   2.613  0.00897 **
## bmi                  -0.175659  0.018546  -9.471 < 2e-16 ***
## log(totchol)          1.027651  0.328788   3.126  0.00177 **
## prevchd              -0.710396  0.232885  -3.050  0.00229 **
## death                0.905432  0.191588   4.726 2.29e-06 ***
## cvd                  0.522970  0.207219   2.524  0.01161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) cage  sex2  edc_b_ bmi   lg(tt) prvchd death
## cage       0.091
## sex2      -0.037 -0.019
## edc_bnryhg_-0.068 -0.033 -0.023
## bmi       -0.128 -0.001  0.095 -0.091
## log(totchl) -0.960 -0.096 -0.119  0.024 -0.135
## prevchd   -0.047 -0.182  0.056  0.012  0.043  0.032
## death     -0.025 -0.194  0.092 -0.057 -0.002  0.004 -0.028
## cvd       0.050 -0.043  0.121 -0.016 -0.073 -0.051 -0.163 -0.339

```

Number of cigarettes ~ age + sex + covariates

systolic bp ~ smoking+ covariates

diastolic bp ~ smoking + covariates

total cholesterol ~ smoking + covariates