Final_Project

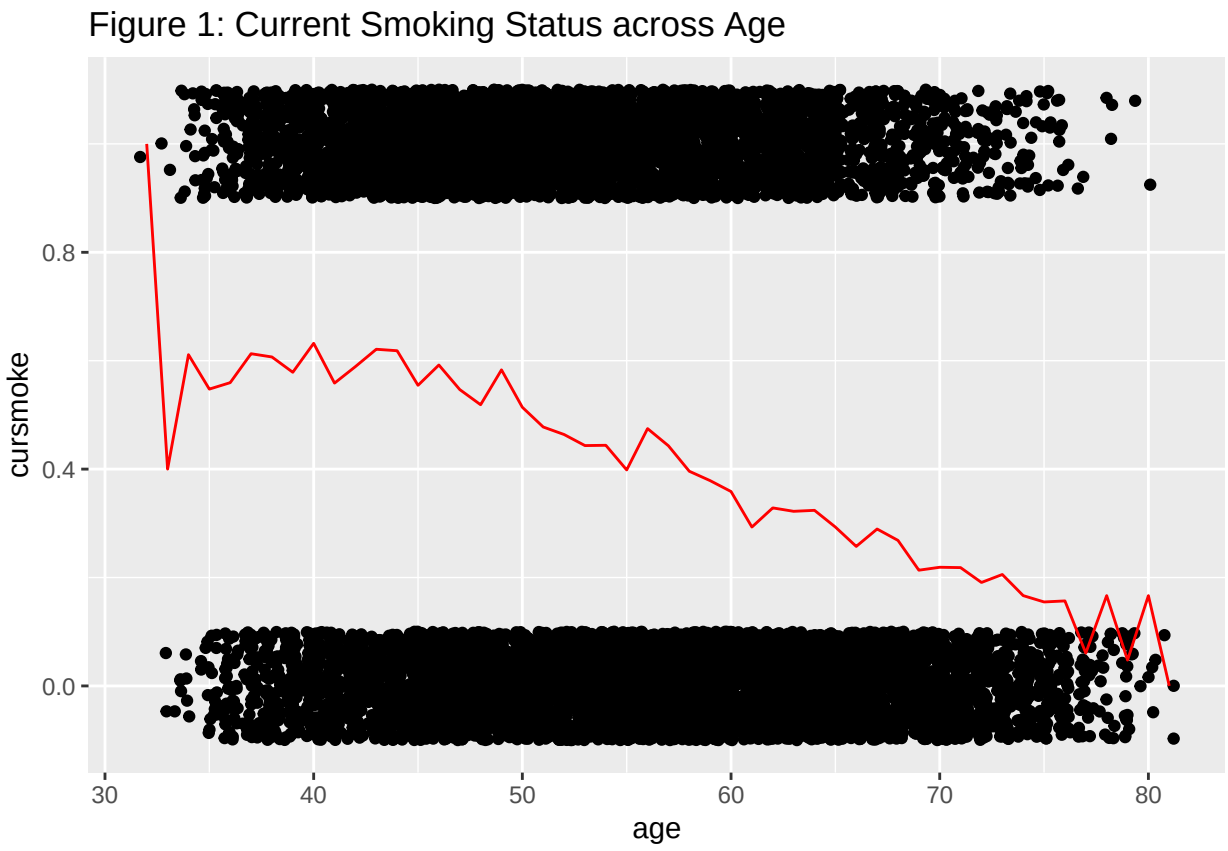
Mengqi Zhu

2018/11/29

```
smoke <- read.csv(file = 'frmgham2.csv') %>%
  clean_names()
```

Figure 1 shows that as individuals age, the likelihood that they are smoking decreases. We can see that when we breaking individuals down by sex, it appears that the overall trend is the same between sexes with males having an overall higher likelihood of being smokers as age increases.

```
smoke %>%
  ggplot(aes(age, cursmoke)) +
  geom_jitter(height = 0.1) +
  stat_summary(fun.y = mean, geom = "line", col = 'red') +
  ggtitle("Figure 1: Current Smoking Status across Age")
```

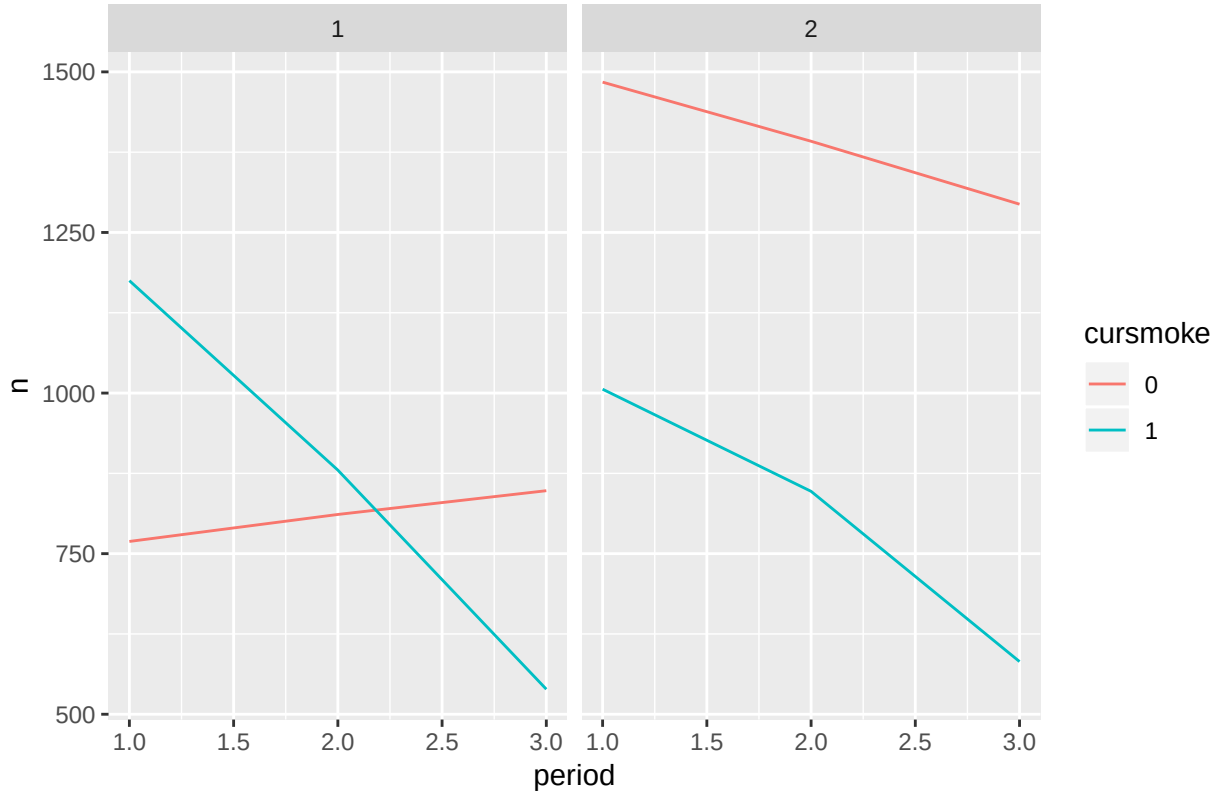



```
##FIGURE OF INTEREST
smoke %>%
  ggplot(aes(age, cursmoke, group = sex)) +
  geom_jitter(height = 0.1, size = 0.5) +
  stat_summary(fun.y = mean, geom = "line", aes(color=paste("mean", sex))) +
  ggtitle("Figure 2: Current Smoking Status across Age") +
  scale_colour_hue(name = "Sex", labels = c("Male", "Female"))
```

Figure 2: Current Smoking Status across Age


```
#longitudinal plot
smoke %>%
  mutate(cursmoke = as.factor(cursmoke),
         sex = as.factor(sex)) %>%
  group_by(period, sex) %>%
  count(cursmoke) %>%
  ggplot(aes(period, n, group = cursmoke, color = cursmoke)) +
  geom_line() +
  facet_wrap(~sex) +
  ggtitle("Figure 3: Current Smoking status across Period by Sex") #period = visit
```

Figure 3: Current Smoking status across Period by Sex



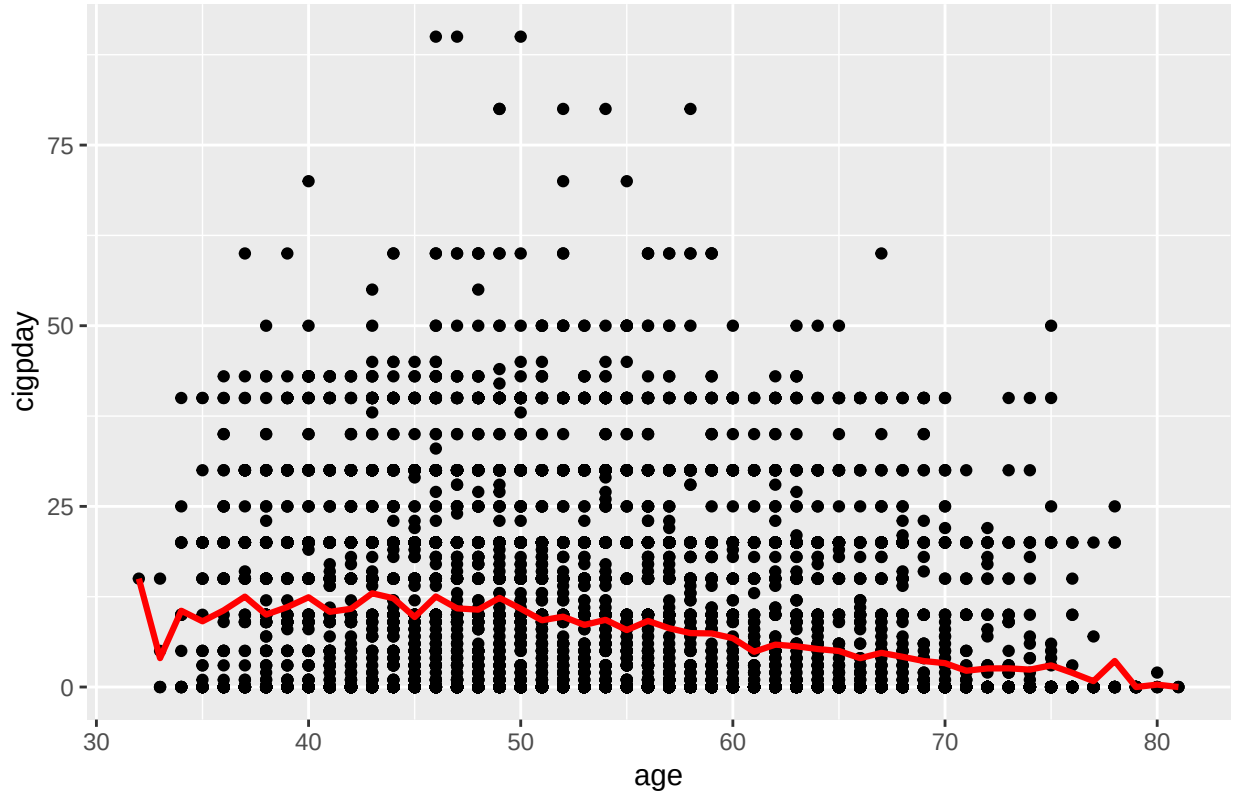
When looking at cigarette packs smoked per day, it appears that the number steadily decreases as individuals get older. The trend once again is the same in each sex however females are smoking less packs a day overall.

```
smoke %>%
  ggplot(aes(age, cigpday)) +
  geom_point() +
  stat_summary(fun.y = mean, geom = "line", size = 1.1, col = 'red') +
  ggtitle("Figure 4: Packs a Day across Age")
```

```
## Warning: Removed 79 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 79 rows containing missing values (geom_point).
```

Figure 4: Packs a Day across Age

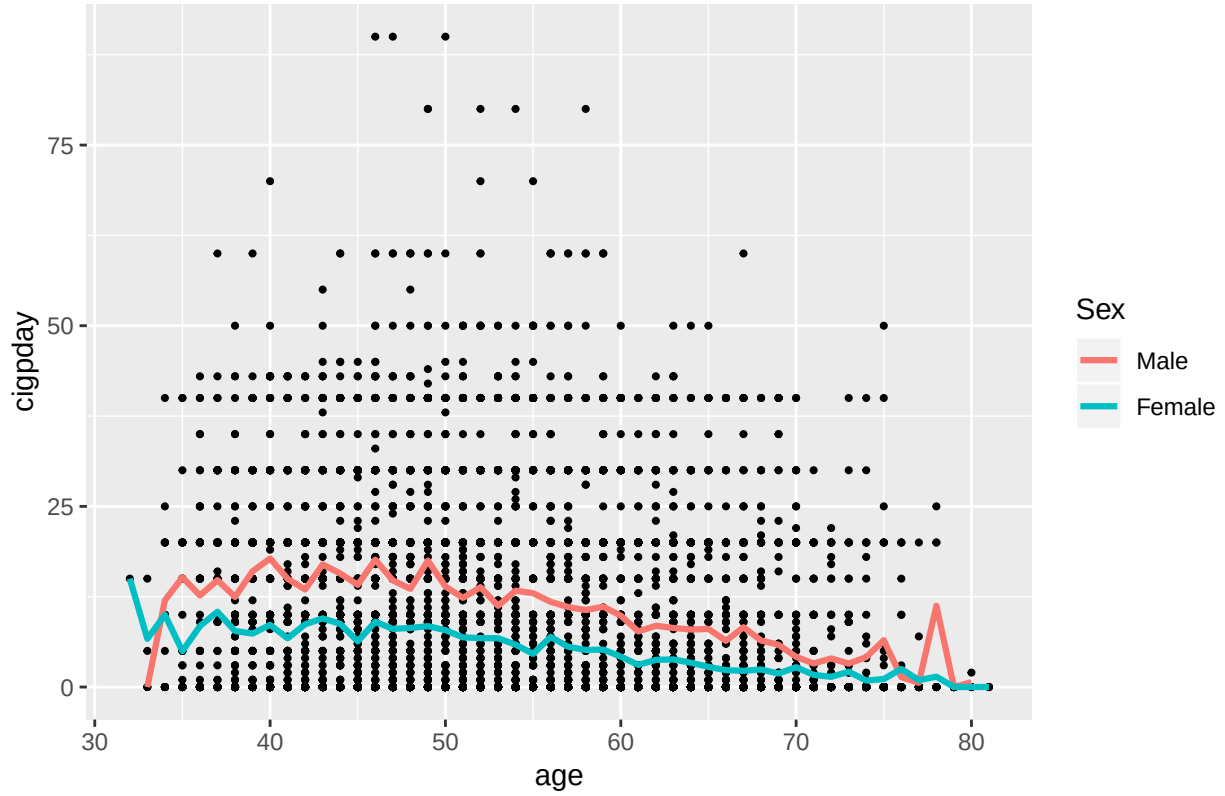



```
smoke %>%
  ggplot(aes(age, cigpday, group = sex)) +
  geom_point(size = 0.75) +
  stat_summary(fun.y = mean, geom = "line", size = 1.1, aes(color=paste("mean", sex))) +
  ggtitle("Figure 5: Cigarettes per Day across Age") +
  scale_colour_hue(name = "Sex", labels = c("Male", "Female"))
```

```
## Warning: Removed 79 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 79 rows containing missing values (geom_point).
```

Figure 5: Cigarettes per Day across Age

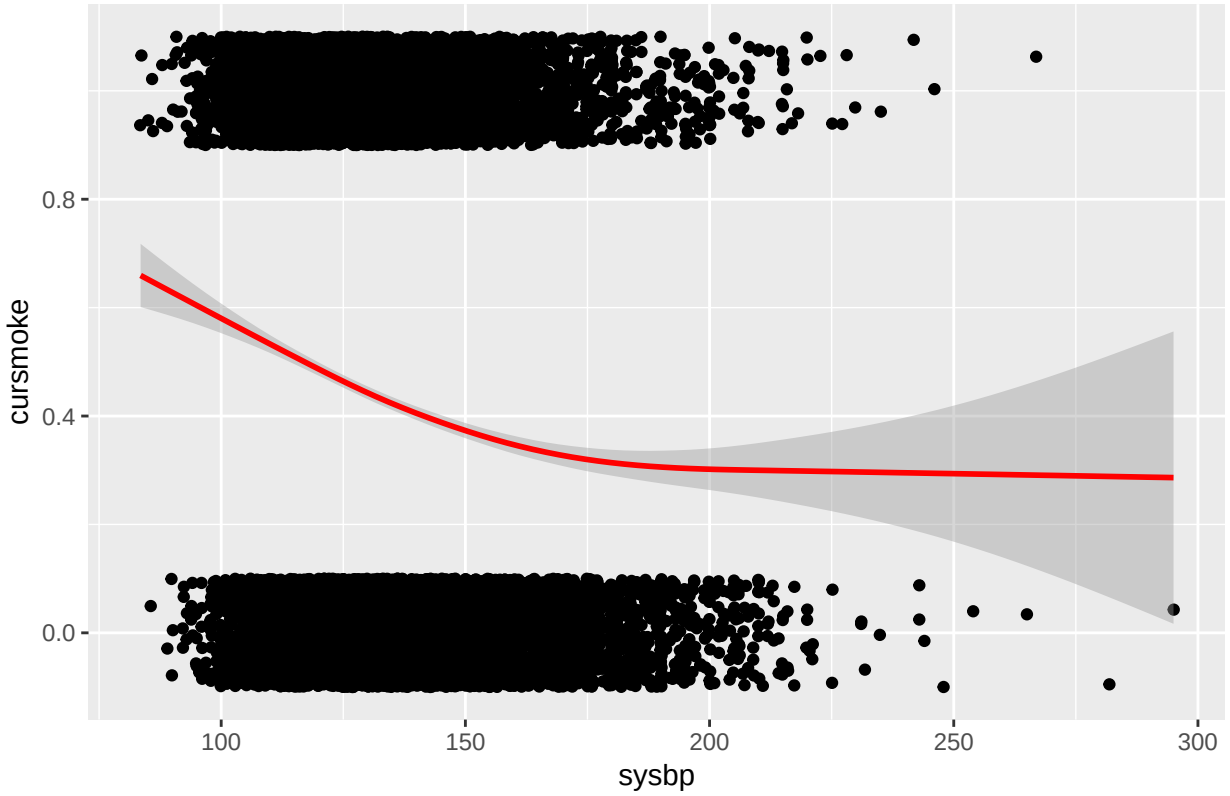


In Figure 5, we see that as systolic blood pressure increases the likelihood of smoking decreases. The trend is not as profound in Figure 6 with diastolic BP or with serum total cholesterol in Figure 7.

```
smoke %>%
  ggplot(aes(sysbp, cursmoke)) +
  geom_jitter(height = 0.1) +
  geom_smooth(color = 'red') +
  ggtitle("Figure 6: Current Smoking Status across Systolic Blood Pressure")
```

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

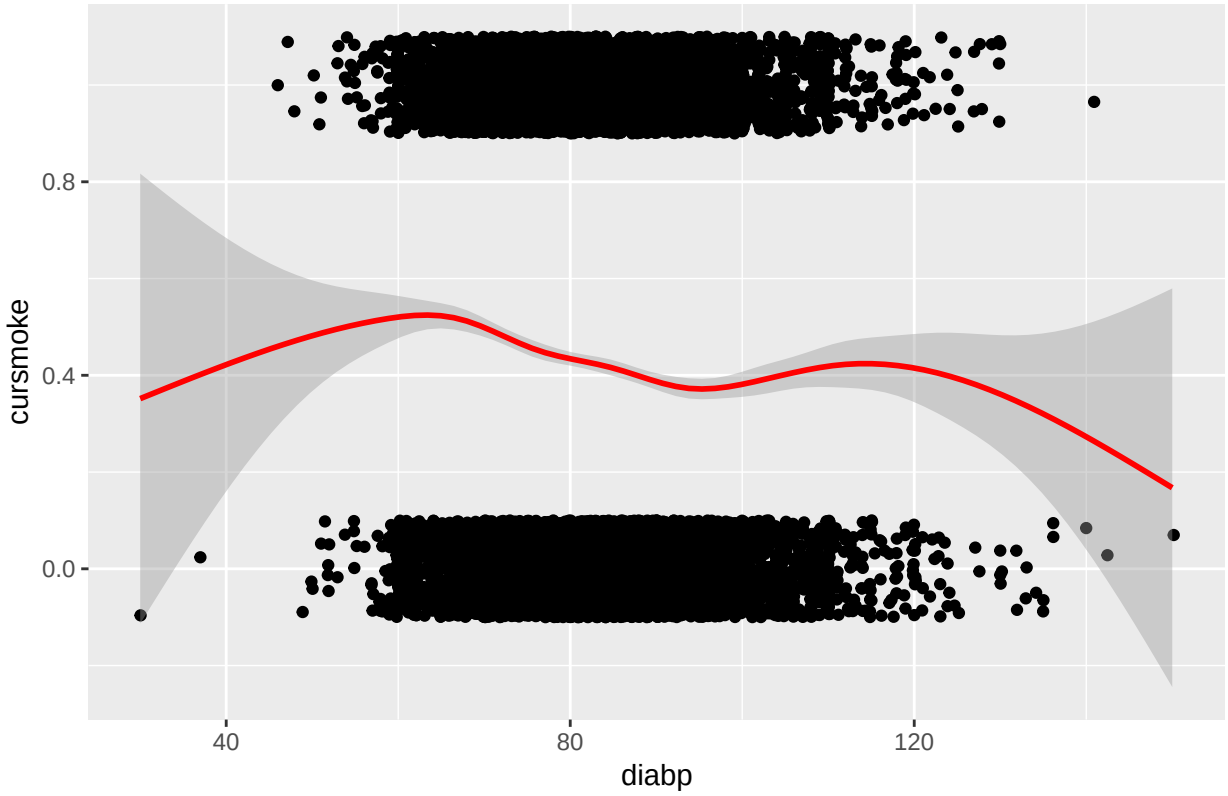
Figure 6: Current Smoking Status across Systolic Blood Pressure




```
smoke %>%
  ggplot(aes(diabp, cursmoke)) +
  geom_jitter(height = 0.1) +
  geom_smooth(color = 'red') +
  ggtitle("Figure 7: Current Smoking Status across Diastolic Blood Pressure")
```

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

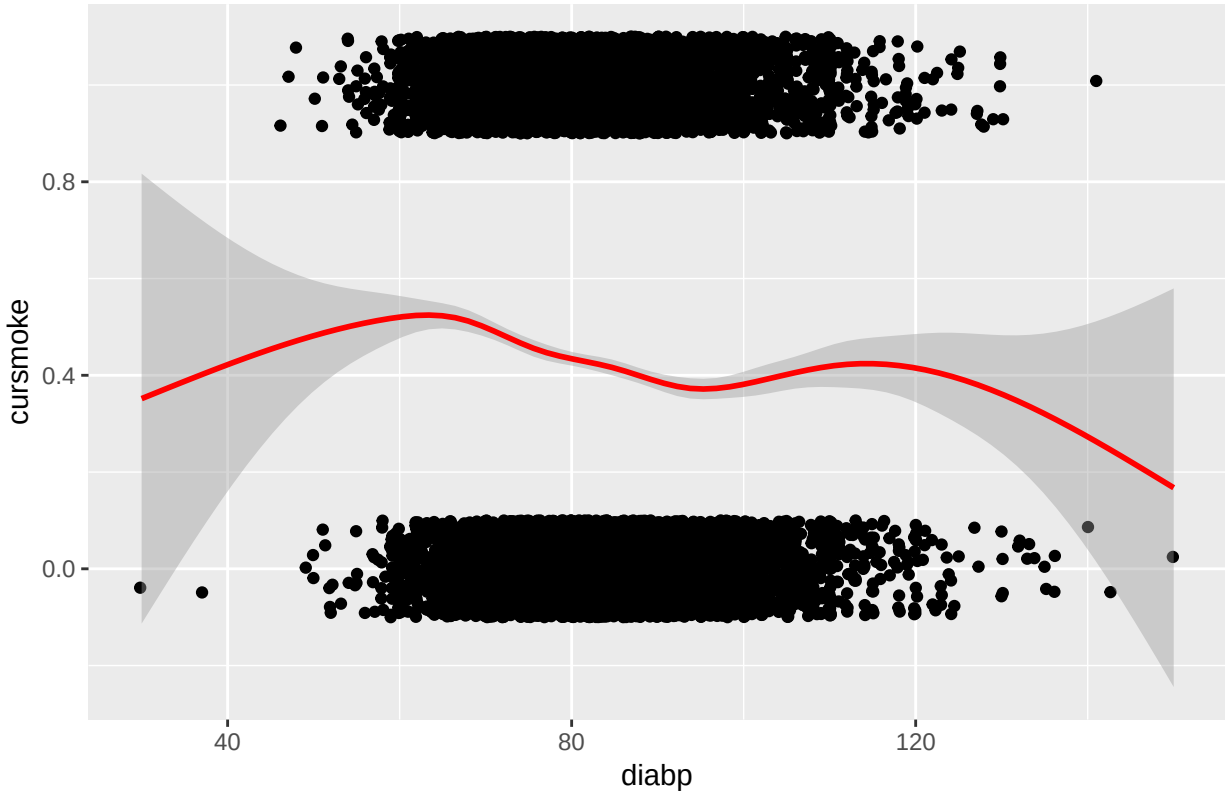
Figure 7: Current Smoking Status across Diastolic Blood Pressure




```
smoke %>%
  ggplot(aes(diabp, cursmoke)) +
  geom_jitter(height = 0.1) +
  geom_smooth(color = 'red') +
  ggtitle("Figure 8: Current Smoking Status across Serum Total Cholestserol")
```

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

Figure 8: Current Smoking Status across Serum Total Cholestserol



Corrplot 1

It appears that the only significant correlation between the variables of interest is between the two BP measurements and systolic BP and age with a value of 0.39. There is negative correlation between age and smoking status, meaning that the smokers appear to be correlated with youger aged individuals.

Corrplot 2

Obvious correlations include BP measurements between bpmeds and prevhyp, age and death, glucose and diabetes, etc. Prevvap was correlated with prevmi and prevchd. Our variable of interest totchol was very highly correlated with ldlc.

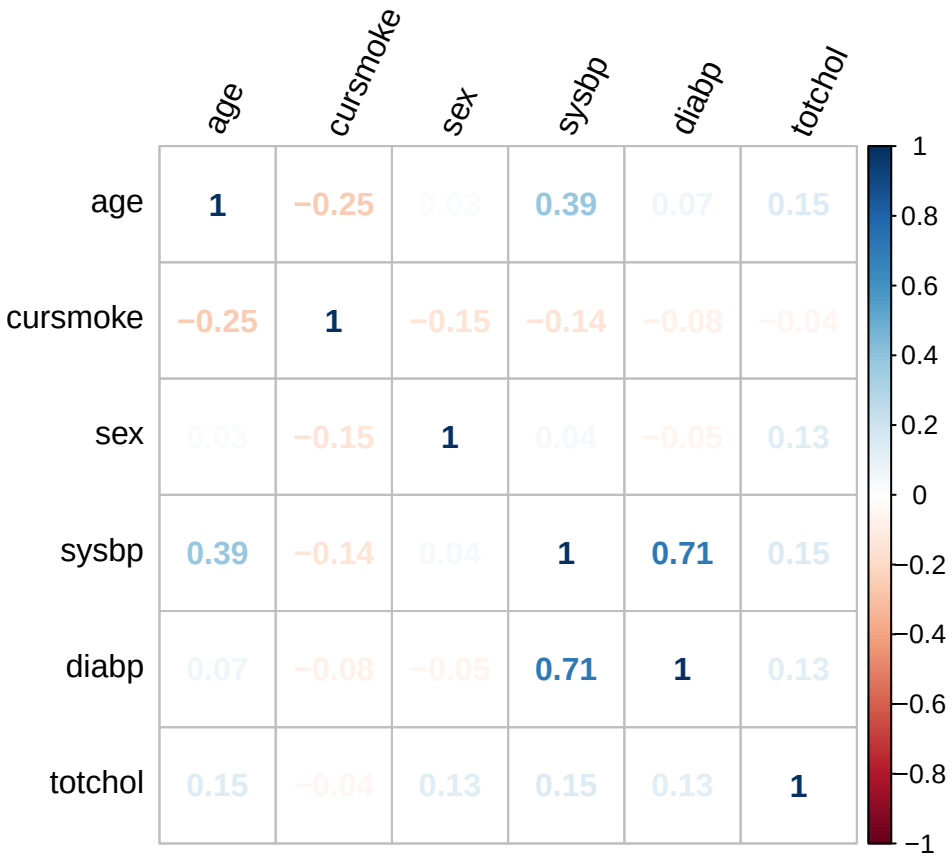
Corrplot3

When evaluating the different follow up conditions, most are highly correlated with each other with the exception of hypertension not being correlated with any other condition and stroke only being correlated to cvd at 0.55. BPs are correlated with hypertension

```
#correlation between all interested variables
corr =  smoke %>%
  dplyr::select(age, cursmoke, sex, sysbp, diabp, totchol) %>%
```

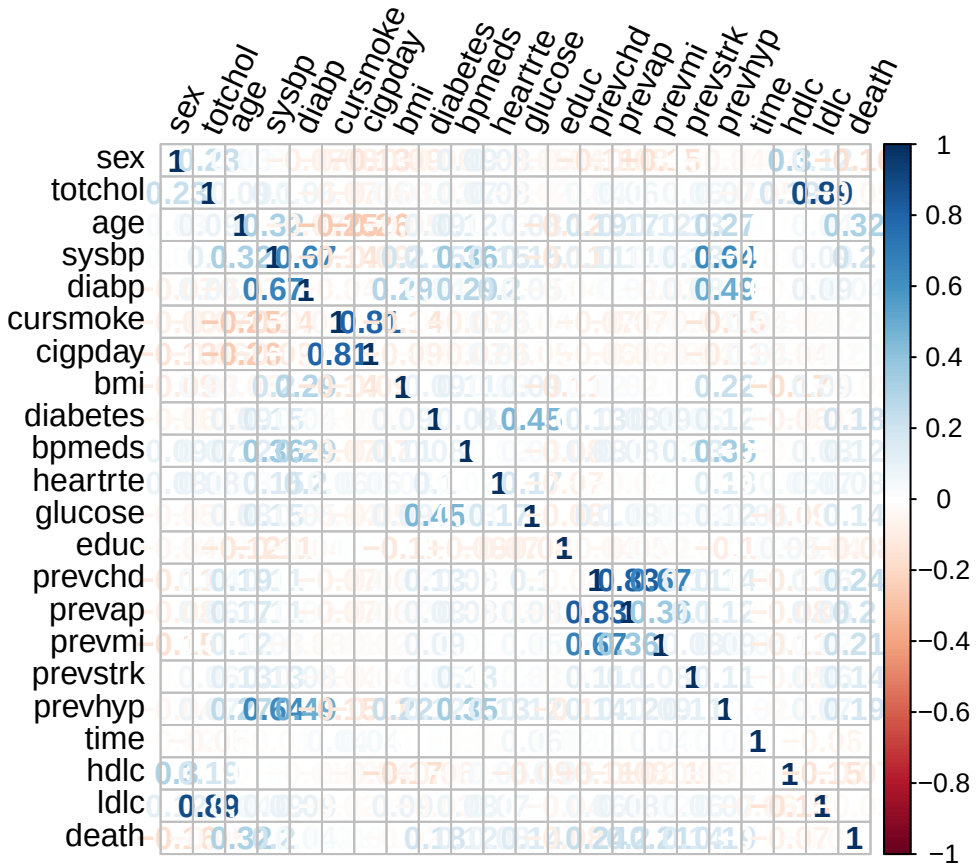
```
  mutate(totchol = as.numeric(totchol)) %>%
  na.omit()

corrplot(cor(corr), method="number",shade.col=NA, tl.col="black", tl.srt=65)
```



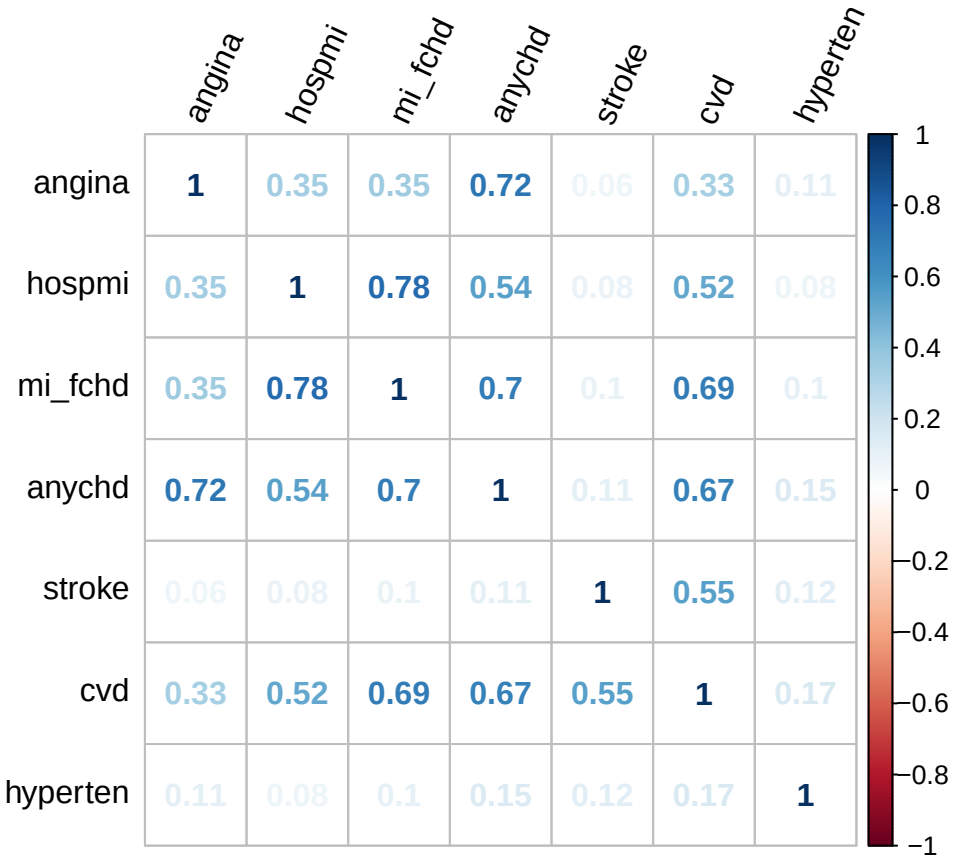

```
corr2 =  smoke %>%
  dplyr::select(-randid, -period,  -25:-39) %>%
  na.omit()

corrplot(cor(corr2), method="number",shade.col=NA, tl.col="black", tl.srt=65)
```




```
corr3 =  smoke %>%
  dplyr::select(25:31) %>%
  na.omit()

corrplot(cor(corr3), method="number",shade.col=NA, tl.col="black", tl.srt=65)
```




```
#checking missing values (5% rule)
pMiss <- function(x){sum(is.na(x))/length(x)*100}
apply(smoke,2,pMiss) #2 indicates columns
```

```
##      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
```

```
#Remove hdlc, ldlc, glucose, and bpmeds for having more than 5% of missing values.
```

Variable Selection and confounder identidication:

```
smoke.vs = smoke %>%
  filter(period==1) %>%
  dplyr::select(c(randid,sex,age,cursmoke,totchol,bmi,heartrte,educ,diabp,sysbp,diabetes,heartrte)) %>%
  mutate(cursmoke = as.factor(cursmoke)) %>%
  na.omit()
```

```r
#(1)
###backward elimination###

back.fit <- glm(cursmoke ~ .-randid, data=smoke.vs,family = 'binomial')
summary(back.fit)
```

```
##
## Call:
## glm(formula = cursmoke ~ . - randid, family = "binomial", data = smoke.vs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4092  -1.0582  -0.4914   1.0580   2.5394
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.5500958  0.4120223  13.470  < 2e-16 ***
## sex         -1.0376486  0.0689654 -15.046  < 2e-16 ***
## age         -0.0510387  0.0044332 -11.513  < 2e-16 ***
## totchol      0.0017787  0.0007837   2.270  0.02324 *
## bmi         -0.0906941  0.0094346  -9.613  < 2e-16 ***
## heartrte     0.0177122  0.0028302   6.258 3.89e-10 ***
## educ        -0.0796102  0.0326062  -2.442  0.01462 *
## diabp       -0.0133757  0.0047094  -2.840  0.00451 **
## sysbp        0.0031624  0.0026610   1.188  0.23467
## diabetes    -0.3741775  0.2136467  -1.751  0.07988 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5893.4  on 4251  degrees of freedom
## Residual deviance: 5343.0  on 4242  degrees of freedom
## AIC: 5363
##
## Number of Fisher Scoring iterations: 4
```

```r
#We then take out the variables with the highest p-value:

step1 <-update(back.fit, . ~ . -sysbp)
summary(step1)
```

```
##
## Call:
## glm(formula = cursmoke ~ sex + age + totchol + bmi + heartrte +
##     educ + diabp + diabetes, family = "binomial", data = smoke.vs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3576  -1.0599  -0.4878   1.0563   2.5386
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.4876224  0.4084683  13.435  < 2e-16 ***
```

```
## sex         -1.0272343  0.0683477 -15.030  < 2e-16 ***
## age         -0.0491518  0.0041328 -11.893  < 2e-16 ***
## totchol      0.0017951  0.0007838   2.290  0.02200 *
## bmi         -0.0905322  0.0094396  -9.591  < 2e-16 ***
## heartrte     0.0179309  0.0028254   6.346 2.21e-10 ***
## educ        -0.0821556  0.0325257  -2.526  0.01154 *
## diabp       -0.0091460  0.0030805  -2.969  0.00299 **
## diabetes    -0.3562231  0.2134563  -1.669  0.09515 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5893.4  on 4251  degrees of freedom
## Residual deviance: 5344.4  on 4243  degrees of freedom
## AIC: 5362.4
##
## Number of Fisher Scoring iterations: 4
```

```
step2 <-update(step1, . ~ . -diabetes)
summary(step2)
```

```
##
## Call:
## glm(formula = cursmoke ~ sex + age + totchol + bmi + heartrte +
##     educ + diabp, family = "binomial", data = smoke.vs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3543  -1.0603  -0.4915   1.0577   2.4120
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.5380902  0.4072657  13.598  < 2e-16 ***
## sex         -1.0238790  0.0682781 -14.996  < 2e-16 ***
## age         -0.0497740  0.0041157 -12.094  < 2e-16 ***
## totchol      0.0017765  0.0007827   2.270  0.02323 *
## bmi         -0.0913411  0.0094227  -9.694  < 2e-16 ***
## heartrte     0.0177088  0.0028207   6.278 3.43e-10 ***
## educ        -0.0818070  0.0325256  -2.515  0.01190 *
## diabp       -0.0090557  0.0030775  -2.943  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5893.4  on 4251  degrees of freedom
## Residual deviance: 5347.2  on 4244  degrees of freedom
## AIC: 5363.2
##
## Number of Fisher Scoring iterations: 4
```

```
#result from backward elimination:
#cursmoke ~ age + sex + totchol + bmi + educ + diabp + heartrte
```

```r
#only backward elimination is recommeded in case of binary outcome
```

If we change the outcome from cursmoke to cigpday:

```r
#(2)
smoke.vs2 = smoke %>%
  filter(period==1) %>%
  dplyr::select(c(randid,sex,age,cigpday,totchol,bmi,heartrte,educ,diabp,sysbp,diabetes,heartrte)) %>%
  na.omit()

###backward elimination###

back.fit2 <- lm(cigpday ~ .-randid, data=smoke.vs2)
summary(back.fit2)
```

```
##
## Call:
## lm(formula = cigpday ~ . - randid, data = smoke.vs2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.286  -7.578  -2.666   6.060  53.966
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.425219   2.006462  17.157  < 2e-16 ***
## sex         -8.380921   0.345296 -24.272  < 2e-16 ***
## age         -0.266460   0.022258 -11.972  < 2e-16 ***
## totchol      0.013848   0.003987   3.473  0.00052 ***
## bmi         -0.297525   0.044917  -6.624 3.94e-11 ***
## heartrte     0.110526   0.014310   7.724 1.40e-14 ***
## educ        -0.459869   0.167694  -2.742  0.00613 **
## diabp       -0.061751   0.023701  -2.605  0.00921 **
## sysbp        0.022666   0.013347   1.698  0.08953 .
## diabetes    -2.027437   1.044535  -1.941  0.05233 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.9 on 4212 degrees of freedom
## Multiple R-squared:  0.1661, Adjusted R-squared:  0.1644
## F-statistic: 93.24 on 9 and 4212 DF,  p-value: < 2.2e-16
```

```r
step1 <-update(back.fit2, . ~ . -sysbp)
summary(step1)
```

```
##
## Call:
## lm(formula = cigpday ~ sex + age + totchol + bmi + heartrte +
##     educ + diabp + diabetes, data = smoke.vs2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.353  -7.665  -2.693   6.036  54.013
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.963493   1.988400  17.081  < 2e-16 ***
## sex         -8.306937   0.342613 -24.246  < 2e-16 ***
## age         -0.252851   0.020770 -12.174  < 2e-16 ***
## totchol      0.013958   0.003988   3.500  0.00047 ***
## bmi         -0.295579   0.044912  -6.581 5.24e-11 ***
## heartrte     0.112123   0.014282   7.850 5.22e-15 ***
## educ        -0.479327   0.167340  -2.864  0.00420 **
## diabp       -0.031455   0.015608  -2.015  0.04393 *
## diabetes    -1.891370   1.041690  -1.816  0.06949 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.91 on 4213 degrees of freedom
## Multiple R-squared:  0.1656, Adjusted R-squared:  0.164
## F-statistic: 104.5 on 8 and 4213 DF,  p-value: < 2.2e-16
```

```
step2 <-update(step1, . ~ . -diabetes)
summary(step2)
```

```
##
## Call:
## lm(formula = cigpday ~ sex + age + totchol + bmi + heartrte +
##     educ + diabp, data = smoke.vs2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.313  -7.658  -2.669   6.034  54.090
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.285393   1.981020  17.307  < 2e-16 ***
## sex         -8.293955   0.342632 -24.207  < 2e-16 ***
## age         -0.256400   0.020683 -12.397  < 2e-16 ***
## totchol      0.013917   0.003989   3.489  0.00049 ***
## bmi         -0.300948   0.044827  -6.714 2.15e-11 ***
## heartrte     0.110859   0.014269   7.769 9.87e-15 ***
## educ        -0.476129   0.167376  -2.845  0.00447 **
## diabp       -0.031201   0.015611  -1.999  0.04571 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.91 on 4214 degrees of freedom
## Multiple R-squared:  0.1649, Adjusted R-squared:  0.1635
## F-statistic: 118.9 on 7 and 4214 DF,  p-value: < 2.2e-16
```

```
#result from backward elimination:
#full model
#cursmoke ~ age + sex + totchol + bmi + educ + diabp + heartrte
```

```
#Vars with missing values for imputation
smoke2 = smoke %>%
  dplyr::select(c(cursmoke,bmi,educ,diabp,age,sex,heartrte,totchol,cigpday))
summary(smoke2)
```

```
##     cursmoke           bmi             educ          diabp
```

```
##  Min.   :0.0000   Min.   :14.43   Min.   :1.00   Min.   : 30.00
##  1st Qu.:0.0000   1st Qu.:23.09   1st Qu.:1.00   1st Qu.: 75.00
##  Median :0.0000   Median :25.48   Median :2.00   Median : 82.00
##  Mean   :0.4325   Mean   :25.88   Mean   :1.99   Mean   : 83.04
##  3rd Qu.:1.0000   3rd Qu.:28.07   3rd Qu.:3.00   3rd Qu.: 90.00
##  Max.   :1.0000   Max.   :56.80   Max.   :4.00   Max.   :150.00
##                   NA's   :52      NA's   :295
##       age             sex           heartrte         totchol
##  Min.   :32.00   Min.   :1.000   Min.   : 37.00   Min.   :107.0
##  1st Qu.:48.00   1st Qu.:1.000   1st Qu.: 69.00   1st Qu.:210.0
##  Median :54.00   Median :2.000   Median : 75.00   Median :238.0
##  Mean   :54.79   Mean   :1.568   Mean   : 76.78   Mean   :241.2
##  3rd Qu.:62.00   3rd Qu.:2.000   3rd Qu.: 85.00   3rd Qu.:268.0
##  Max.   :81.00   Max.   :2.000   Max.   :220.00   Max.   :696.0
##                                  NA's   :6        NA's   :409
##     cigpday
##  Min.   : 0.00
##  1st Qu.: 0.00
##  Median : 0.00
##  Mean   : 8.25
##  3rd Qu.:20.00
##  Max.   :90.00
##  NA's   :79
```

```
#imputation using predictive mean matching
imputed.smoke <- mice(smoke2,m=5,maxit=50,meth='pmm',seed=500)
```

```
##
##  iter imp variable
##   1   1  bmi  educ  heartrte  totchol  cigpday
##   1   2  bmi  educ  heartrte  totchol  cigpday
##   1   3  bmi  educ  heartrte  totchol  cigpday
##   1   4  bmi  educ  heartrte  totchol  cigpday
##   1   5  bmi  educ  heartrte  totchol  cigpday
##   2   1  bmi  educ  heartrte  totchol  cigpday
##   2   2  bmi  educ  heartrte  totchol  cigpday
##   2   3  bmi  educ  heartrte  totchol  cigpday
##   2   4  bmi  educ  heartrte  totchol  cigpday
##   2   5  bmi  educ  heartrte  totchol  cigpday
##   3   1  bmi  educ  heartrte  totchol  cigpday
##   3   2  bmi  educ  heartrte  totchol  cigpday
##   3   3  bmi  educ  heartrte  totchol  cigpday
##   3   4  bmi  educ  heartrte  totchol  cigpday
##   3   5  bmi  educ  heartrte  totchol  cigpday
##   4   1  bmi  educ  heartrte  totchol  cigpday
##   4   2  bmi  educ  heartrte  totchol  cigpday
##   4   3  bmi  educ  heartrte  totchol  cigpday
##   4   4  bmi  educ  heartrte  totchol  cigpday
##   4   5  bmi  educ  heartrte  totchol  cigpday
##   5   1  bmi  educ  heartrte  totchol  cigpday
##   5   2  bmi  educ  heartrte  totchol  cigpday
##   5   3  bmi  educ  heartrte  totchol  cigpday
##   5   4  bmi  educ  heartrte  totchol  cigpday
##   5   5  bmi  educ  heartrte  totchol  cigpday
##   6   1  bmi  educ  heartrte  totchol  cigpday
```

```
##   6   2  bmi  educ  heartrte  totchol  cigpday
##   6   3  bmi  educ  heartrte  totchol  cigpday
##   6   4  bmi  educ  heartrte  totchol  cigpday
##   6   5  bmi  educ  heartrte  totchol  cigpday
##   7   1  bmi  educ  heartrte  totchol  cigpday
##   7   2  bmi  educ  heartrte  totchol  cigpday
##   7   3  bmi  educ  heartrte  totchol  cigpday
##   7   4  bmi  educ  heartrte  totchol  cigpday
##   7   5  bmi  educ  heartrte  totchol  cigpday
##   8   1  bmi  educ  heartrte  totchol  cigpday
##   8   2  bmi  educ  heartrte  totchol  cigpday
##   8   3  bmi  educ  heartrte  totchol  cigpday
##   8   4  bmi  educ  heartrte  totchol  cigpday
##   8   5  bmi  educ  heartrte  totchol  cigpday
##   9   1  bmi  educ  heartrte  totchol  cigpday
##   9   2  bmi  educ  heartrte  totchol  cigpday
##   9   3  bmi  educ  heartrte  totchol  cigpday
##   9   4  bmi  educ  heartrte  totchol  cigpday
##   9   5  bmi  educ  heartrte  totchol  cigpday
##  10   1  bmi  educ  heartrte  totchol  cigpday
##  10   2  bmi  educ  heartrte  totchol  cigpday
##  10   3  bmi  educ  heartrte  totchol  cigpday
##  10   4  bmi  educ  heartrte  totchol  cigpday
##  10   5  bmi  educ  heartrte  totchol  cigpday
##  11   1  bmi  educ  heartrte  totchol  cigpday
##  11   2  bmi  educ  heartrte  totchol  cigpday
##  11   3  bmi  educ  heartrte  totchol  cigpday
##  11   4  bmi  educ  heartrte  totchol  cigpday
##  11   5  bmi  educ  heartrte  totchol  cigpday
##  12   1  bmi  educ  heartrte  totchol  cigpday
##  12   2  bmi  educ  heartrte  totchol  cigpday
##  12   3  bmi  educ  heartrte  totchol  cigpday
##  12   4  bmi  educ  heartrte  totchol  cigpday
##  12   5  bmi  educ  heartrte  totchol  cigpday
##  13   1  bmi  educ  heartrte  totchol  cigpday
##  13   2  bmi  educ  heartrte  totchol  cigpday
##  13   3  bmi  educ  heartrte  totchol  cigpday
##  13   4  bmi  educ  heartrte  totchol  cigpday
##  13   5  bmi  educ  heartrte  totchol  cigpday
##  14   1  bmi  educ  heartrte  totchol  cigpday
##  14   2  bmi  educ  heartrte  totchol  cigpday
##  14   3  bmi  educ  heartrte  totchol  cigpday
##  14   4  bmi  educ  heartrte  totchol  cigpday
##  14   5  bmi  educ  heartrte  totchol  cigpday
##  15   1  bmi  educ  heartrte  totchol  cigpday
##  15   2  bmi  educ  heartrte  totchol  cigpday
##  15   3  bmi  educ  heartrte  totchol  cigpday
##  15   4  bmi  educ  heartrte  totchol  cigpday
##  15   5  bmi  educ  heartrte  totchol  cigpday
##  16   1  bmi  educ  heartrte  totchol  cigpday
##  16   2  bmi  educ  heartrte  totchol  cigpday
##  16   3  bmi  educ  heartrte  totchol  cigpday
##  16   4  bmi  educ  heartrte  totchol  cigpday
##  16   5  bmi  educ  heartrte  totchol  cigpday
```

```
##   17   1  bmi  educ  heartrte  totchol  cigpday
##   17   2  bmi  educ  heartrte  totchol  cigpday
##   17   3  bmi  educ  heartrte  totchol  cigpday
##   17   4  bmi  educ  heartrte  totchol  cigpday
##   17   5  bmi  educ  heartrte  totchol  cigpday
##   18   1  bmi  educ  heartrte  totchol  cigpday
##   18   2  bmi  educ  heartrte  totchol  cigpday
##   18   3  bmi  educ  heartrte  totchol  cigpday
##   18   4  bmi  educ  heartrte  totchol  cigpday
##   18   5  bmi  educ  heartrte  totchol  cigpday
##   19   1  bmi  educ  heartrte  totchol  cigpday
##   19   2  bmi  educ  heartrte  totchol  cigpday
##   19   3  bmi  educ  heartrte  totchol  cigpday
##   19   4  bmi  educ  heartrte  totchol  cigpday
##   19   5  bmi  educ  heartrte  totchol  cigpday
##   20   1  bmi  educ  heartrte  totchol  cigpday
##   20   2  bmi  educ  heartrte  totchol  cigpday
##   20   3  bmi  educ  heartrte  totchol  cigpday
##   20   4  bmi  educ  heartrte  totchol  cigpday
##   20   5  bmi  educ  heartrte  totchol  cigpday
##   21   1  bmi  educ  heartrte  totchol  cigpday
##   21   2  bmi  educ  heartrte  totchol  cigpday
##   21   3  bmi  educ  heartrte  totchol  cigpday
##   21   4  bmi  educ  heartrte  totchol  cigpday
##   21   5  bmi  educ  heartrte  totchol  cigpday
##   22   1  bmi  educ  heartrte  totchol  cigpday
##   22   2  bmi  educ  heartrte  totchol  cigpday
##   22   3  bmi  educ  heartrte  totchol  cigpday
##   22   4  bmi  educ  heartrte  totchol  cigpday
##   22   5  bmi  educ  heartrte  totchol  cigpday
##   23   1  bmi  educ  heartrte  totchol  cigpday
##   23   2  bmi  educ  heartrte  totchol  cigpday
##   23   3  bmi  educ  heartrte  totchol  cigpday
##   23   4  bmi  educ  heartrte  totchol  cigpday
##   23   5  bmi  educ  heartrte  totchol  cigpday
##   24   1  bmi  educ  heartrte  totchol  cigpday
##   24   2  bmi  educ  heartrte  totchol  cigpday
##   24   3  bmi  educ  heartrte  totchol  cigpday
##   24   4  bmi  educ  heartrte  totchol  cigpday
##   24   5  bmi  educ  heartrte  totchol  cigpday
##   25   1  bmi  educ  heartrte  totchol  cigpday
##   25   2  bmi  educ  heartrte  totchol  cigpday
##   25   3  bmi  educ  heartrte  totchol  cigpday
##   25   4  bmi  educ  heartrte  totchol  cigpday
##   25   5  bmi  educ  heartrte  totchol  cigpday
##   26   1  bmi  educ  heartrte  totchol  cigpday
##   26   2  bmi  educ  heartrte  totchol  cigpday
##   26   3  bmi  educ  heartrte  totchol  cigpday
##   26   4  bmi  educ  heartrte  totchol  cigpday
##   26   5  bmi  educ  heartrte  totchol  cigpday
##   27   1  bmi  educ  heartrte  totchol  cigpday
##   27   2  bmi  educ  heartrte  totchol  cigpday
##   27   3  bmi  educ  heartrte  totchol  cigpday
##   27   4  bmi  educ  heartrte  totchol  cigpday
```

```
##   27   5  bmi  educ  heartrte  totchol  cigpday
##   28   1  bmi  educ  heartrte  totchol  cigpday
##   28   2  bmi  educ  heartrte  totchol  cigpday
##   28   3  bmi  educ  heartrte  totchol  cigpday
##   28   4  bmi  educ  heartrte  totchol  cigpday
##   28   5  bmi  educ  heartrte  totchol  cigpday
##   29   1  bmi  educ  heartrte  totchol  cigpday
##   29   2  bmi  educ  heartrte  totchol  cigpday
##   29   3  bmi  educ  heartrte  totchol  cigpday
##   29   4  bmi  educ  heartrte  totchol  cigpday
##   29   5  bmi  educ  heartrte  totchol  cigpday
##   30   1  bmi  educ  heartrte  totchol  cigpday
##   30   2  bmi  educ  heartrte  totchol  cigpday
##   30   3  bmi  educ  heartrte  totchol  cigpday
##   30   4  bmi  educ  heartrte  totchol  cigpday
##   30   5  bmi  educ  heartrte  totchol  cigpday
##   31   1  bmi  educ  heartrte  totchol  cigpday
##   31   2  bmi  educ  heartrte  totchol  cigpday
##   31   3  bmi  educ  heartrte  totchol  cigpday
##   31   4  bmi  educ  heartrte  totchol  cigpday
##   31   5  bmi  educ  heartrte  totchol  cigpday
##   32   1  bmi  educ  heartrte  totchol  cigpday
##   32   2  bmi  educ  heartrte  totchol  cigpday
##   32   3  bmi  educ  heartrte  totchol  cigpday
##   32   4  bmi  educ  heartrte  totchol  cigpday
##   32   5  bmi  educ  heartrte  totchol  cigpday
##   33   1  bmi  educ  heartrte  totchol  cigpday
##   33   2  bmi  educ  heartrte  totchol  cigpday
##   33   3  bmi  educ  heartrte  totchol  cigpday
##   33   4  bmi  educ  heartrte  totchol  cigpday
##   33   5  bmi  educ  heartrte  totchol  cigpday
##   34   1  bmi  educ  heartrte  totchol  cigpday
##   34   2  bmi  educ  heartrte  totchol  cigpday
##   34   3  bmi  educ  heartrte  totchol  cigpday
##   34   4  bmi  educ  heartrte  totchol  cigpday
##   34   5  bmi  educ  heartrte  totchol  cigpday
##   35   1  bmi  educ  heartrte  totchol  cigpday
##   35   2  bmi  educ  heartrte  totchol  cigpday
##   35   3  bmi  educ  heartrte  totchol  cigpday
##   35   4  bmi  educ  heartrte  totchol  cigpday
##   35   5  bmi  educ  heartrte  totchol  cigpday
##   36   1  bmi  educ  heartrte  totchol  cigpday
##   36   2  bmi  educ  heartrte  totchol  cigpday
##   36   3  bmi  educ  heartrte  totchol  cigpday
##   36   4  bmi  educ  heartrte  totchol  cigpday
##   36   5  bmi  educ  heartrte  totchol  cigpday
##   37   1  bmi  educ  heartrte  totchol  cigpday
##   37   2  bmi  educ  heartrte  totchol  cigpday
##   37   3  bmi  educ  heartrte  totchol  cigpday
##   37   4  bmi  educ  heartrte  totchol  cigpday
##   37   5  bmi  educ  heartrte  totchol  cigpday
##   38   1  bmi  educ  heartrte  totchol  cigpday
##   38   2  bmi  educ  heartrte  totchol  cigpday
##   38   3  bmi  educ  heartrte  totchol  cigpday
```

```
##   38   4  bmi  educ  heartrte  totchol  cigpday
##   38   5  bmi  educ  heartrte  totchol  cigpday
##   39   1  bmi  educ  heartrte  totchol  cigpday
##   39   2  bmi  educ  heartrte  totchol  cigpday
##   39   3  bmi  educ  heartrte  totchol  cigpday
##   39   4  bmi  educ  heartrte  totchol  cigpday
##   39   5  bmi  educ  heartrte  totchol  cigpday
##   40   1  bmi  educ  heartrte  totchol  cigpday
##   40   2  bmi  educ  heartrte  totchol  cigpday
##   40   3  bmi  educ  heartrte  totchol  cigpday
##   40   4  bmi  educ  heartrte  totchol  cigpday
##   40   5  bmi  educ  heartrte  totchol  cigpday
##   41   1  bmi  educ  heartrte  totchol  cigpday
##   41   2  bmi  educ  heartrte  totchol  cigpday
##   41   3  bmi  educ  heartrte  totchol  cigpday
##   41   4  bmi  educ  heartrte  totchol  cigpday
##   41   5  bmi  educ  heartrte  totchol  cigpday
##   42   1  bmi  educ  heartrte  totchol  cigpday
##   42   2  bmi  educ  heartrte  totchol  cigpday
##   42   3  bmi  educ  heartrte  totchol  cigpday
##   42   4  bmi  educ  heartrte  totchol  cigpday
##   42   5  bmi  educ  heartrte  totchol  cigpday
##   43   1  bmi  educ  heartrte  totchol  cigpday
##   43   2  bmi  educ  heartrte  totchol  cigpday
##   43   3  bmi  educ  heartrte  totchol  cigpday
##   43   4  bmi  educ  heartrte  totchol  cigpday
##   43   5  bmi  educ  heartrte  totchol  cigpday
##   44   1  bmi  educ  heartrte  totchol  cigpday
##   44   2  bmi  educ  heartrte  totchol  cigpday
##   44   3  bmi  educ  heartrte  totchol  cigpday
##   44   4  bmi  educ  heartrte  totchol  cigpday
##   44   5  bmi  educ  heartrte  totchol  cigpday
##   45   1  bmi  educ  heartrte  totchol  cigpday
##   45   2  bmi  educ  heartrte  totchol  cigpday
##   45   3  bmi  educ  heartrte  totchol  cigpday
##   45   4  bmi  educ  heartrte  totchol  cigpday
##   45   5  bmi  educ  heartrte  totchol  cigpday
##   46   1  bmi  educ  heartrte  totchol  cigpday
##   46   2  bmi  educ  heartrte  totchol  cigpday
##   46   3  bmi  educ  heartrte  totchol  cigpday
##   46   4  bmi  educ  heartrte  totchol  cigpday
##   46   5  bmi  educ  heartrte  totchol  cigpday
##   47   1  bmi  educ  heartrte  totchol  cigpday
##   47   2  bmi  educ  heartrte  totchol  cigpday
##   47   3  bmi  educ  heartrte  totchol  cigpday
##   47   4  bmi  educ  heartrte  totchol  cigpday
##   47   5  bmi  educ  heartrte  totchol  cigpday
##   48   1  bmi  educ  heartrte  totchol  cigpday
##   48   2  bmi  educ  heartrte  totchol  cigpday
##   48   3  bmi  educ  heartrte  totchol  cigpday
##   48   4  bmi  educ  heartrte  totchol  cigpday
##   48   5  bmi  educ  heartrte  totchol  cigpday
##   49   1  bmi  educ  heartrte  totchol  cigpday
##   49   2  bmi  educ  heartrte  totchol  cigpday
```

```
##   49   3  bmi  educ  heartrte  totchol  cigpday
##   49   4  bmi  educ  heartrte  totchol  cigpday
##   49   5  bmi  educ  heartrte  totchol  cigpday
##   50   1  bmi  educ  heartrte  totchol  cigpday
##   50   2  bmi  educ  heartrte  totchol  cigpday
##   50   3  bmi  educ  heartrte  totchol  cigpday
##   50   4  bmi  educ  heartrte  totchol  cigpday
##   50   5  bmi  educ  heartrte  totchol  cigpday
```

```
#summary(imputed.smoke)

#return to completed dataset
completedData <- complete(imputed.smoke,1)
completedData$randid = smoke$randid
summary(completedData)
```

```
##     cursmoke           bmi             educ           diabp
##  Min.   :0.0000   Min.   :14.43   Min.   :1.000   Min.   : 30.00
##  1st Qu.:0.0000   1st Qu.:23.09   1st Qu.:1.000   1st Qu.: 75.00
##  Median :0.0000   Median :25.48   Median :2.000   Median : 82.00
##  Mean   :0.4325   Mean   :25.88   Mean   :1.991   Mean   : 83.04
##  3rd Qu.:1.0000   3rd Qu.:28.07   3rd Qu.:3.000   3rd Qu.: 90.00
##  Max.   :1.0000   Max.   :56.80   Max.   :4.000   Max.   :150.00
##       age              sex           heartrte        totchol
##  Min.   :32.00   Min.   :1.000   Min.   : 37.00   Min.   :107.0
##  1st Qu.:48.00   1st Qu.:1.000   1st Qu.: 69.00   1st Qu.:210.0
##  Median :54.00   Median :2.000   Median : 75.00   Median :239.0
##  Mean   :54.79   Mean   :1.568   Mean   : 76.78   Mean   :241.4
##  3rd Qu.:62.00   3rd Qu.:2.000   3rd Qu.: 85.00   3rd Qu.:268.0
##  Max.   :81.00   Max.   :2.000   Max.   :220.00   Max.   :696.0
##     cigpday           randid
##  Min.   : 0.000   Min.   :   2448
##  1st Qu.: 0.000   1st Qu.:2474378
##  Median : 0.000   Median :5006008
##  Mean   : 8.305   Mean   :5004741
##  3rd Qu.:20.000   3rd Qu.:7472730
##  Max.   :90.000   Max.   :9999312
```