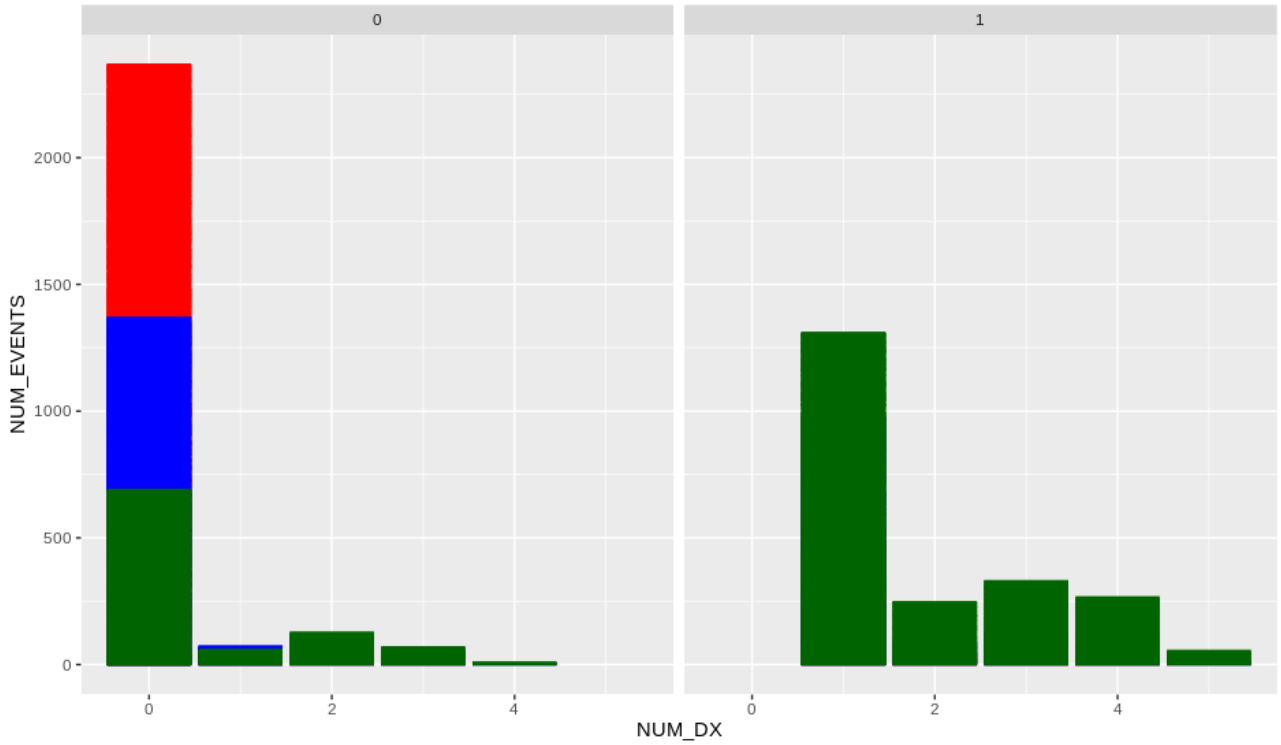
```
#Summary Graphs: # of comorbidities & events by period for those
#with hypertension (1) and without (0)
ggplot() +
  geom_col(data = fr_sum1, aes(x = NUM_DX, y = NUM_EVENTS), color="red") +
  geom_col(data = fr_sum2, aes(x = NUM_DX, y = NUM_EVENTS), color="blue") +
  geom_col(data = fr_sum3, aes(x = NUM_DX, y = NUM_EVENTS), color="dark green") +
  facet_wrap(~PREVHYP)
  labs(title = "Summary: Number of Events x Number of Diagnoses, At Risk Population", x = "Number of Diagnoses",
y = "Number of Events")
```

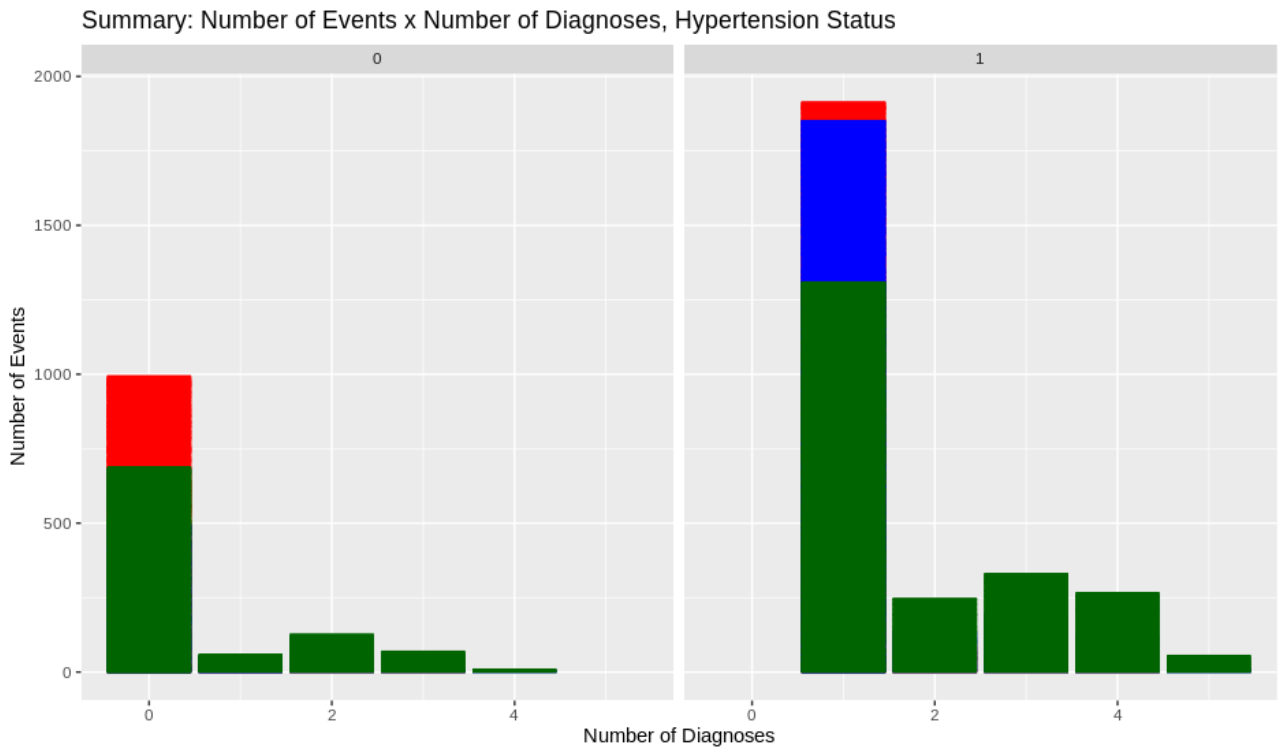


Those with PREVHYP all had at least 1 pre-existing condition (Hypertension). However, note how there were NO pre-existing conditions in Period 1 (Red) and few in Period 2 (Blue)

Here is the same graph again with Period 1 and 2 recoded for high SYS and DIA BP readings:

```
  fr_sum1a <- fr_ex %>% filter(PERIOD==1)
  fr_sum1a <- fr_sum1a %>% mutate(PREVHYP=PREVHYP + as.numeric(PREVHYP==0 & HTNGROUP!="Low" & HTNGROUP!="Normal"))
  fr_sum1a <- fr_sum1a %>% mutate(NUM_DX = PREVAP + PREVCHD + PREVMI + PREVSTRK + DIABETES + PREVHYP)
  fr_sum1a <- fr_sum1a %>% mutate(NUM_EVENTS = ANGINA + HOSPMI + MI_FCHD + STROKE + CVD + HYPERTEN)


  fr_sum2a <- fr_ex %>% filter(PERIOD==2)
  fr_sum2a <- fr_sum2a %>% mutate(PREVHYP=PREVHYP + as.numeric(PREVHYP==0 & HTNGROUP!="Low" & HTNGROUP!="Normal"))
  fr_sum2a <- fr_sum2a %>% mutate(NUM_DX = PREVAP + PREVCHD + PREVMI + PREVSTRK + DIABETES + PREVHYP)
  fr_sum2a <- fr_sum2a %>% mutate(NUM_EVENTS = ANGINA + HOSPMI + MI_FCHD + STROKE + CVD + HYPERTEN)


  # Summary Regraph
  ggplot() +
    geom_col(data = fr_sum1a, aes(x = NUM_DX, y = NUM_EVENTS), color="red") +
```

```
    geom_col(data = fr_sum2a, aes(x = NUM_DX, y = NUM_EVENTS), color="blue") +
    geom_col(data = fr_sum3, aes(x = NUM_DX, y = NUM_EVENTS), color="dark green") +
    facet_wrap(~PREVHYP) +
  labs(title = "Summary: Number of Events x Number of Diagnoses, Hypertension Status", x = "Number of Diagnoses",
y = "Number of Events")
```

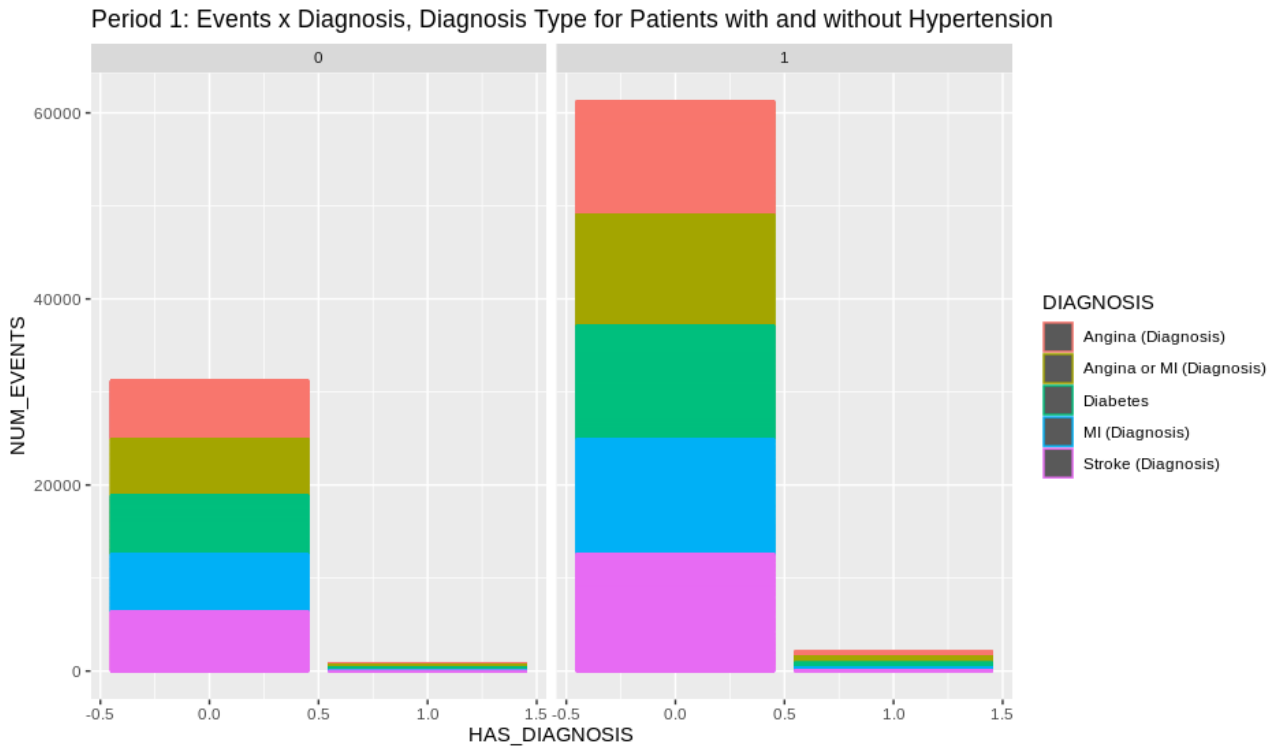


Based on BP READINGS, Many more individuals in Periods 1 and 2 had Hypertension than had an official diagnosis of same.

With our newly coded data, we can now take a look at WHAT diagnoses and events patients had in each period.

```
# Keep recoded diagnosis data
fr_sum1 <- fr_sum1a
fr_sum2 <- fr_sum2a

# Period 1 - Gather diagnoses
fr_gathered1 <- fr_sum1 %>% gather(DIAGNOSIS, HAS_DIAGNOSIS, c(PREVAP, PREVCHD, PREVMI, PREVSTRK, DIABETES))

# Recode to readable
fr_gathered1 <- fr_gathered1 %>% mutate(DIAGNOSIS = recode(DIAGNOSIS,
                                        PREVAP = "Angina (Diagnosis)",
                                        PREVCHD = "Angina or MI (Diagnosis)",
                                        PREVMI = "MI (Diagnosis)",
                                        PREVSTRK = "Stroke (Diagnosis)",
                                        DIABETES = "Diabetes"))
# Period 1 - Gather events
```

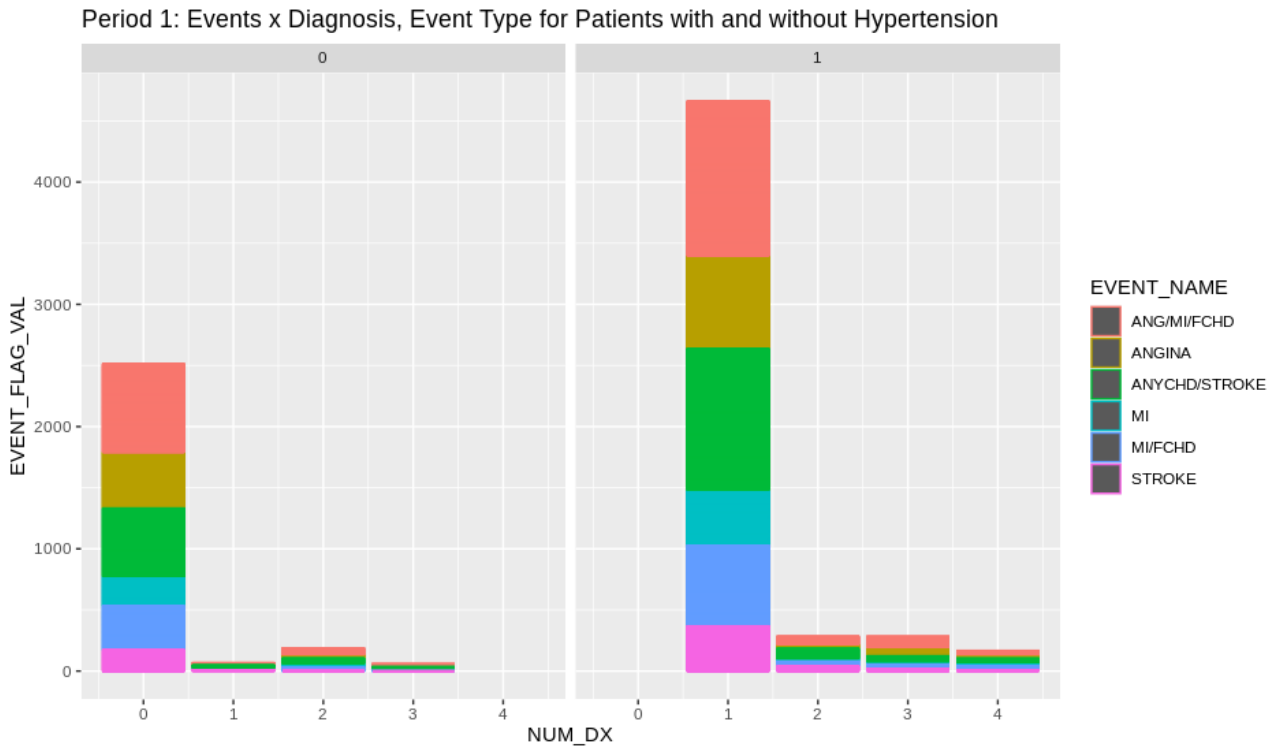
```
fr_gathered1 <- fr_gathered1 %>% gather(EVENT_NAME, EVENT_FLAG_VAL, c(ANGINA, HOSPMI, MI_FCHD, ANYCHD, STROKE,
CVD))
fr_gathered1 <- fr_gathered1 %>% mutate(EVENT_NAME = recode(EVENT_NAME,
                                        ANGINA = "ANGINA",
                                        HOSPMI = "MI",
                                        MI_FCHD = "MI/FCHD",
                                        ANYCHD = "ANG/MI/FCHD",
                                        STROKE = "STROKE",
                                        CVD = "ANYCHD/STROKE"))

#Period 1 - Number of Events by Diagnoses, Diagnosis Type; Facet Grid: PREVHYP
  ggplot(fr_gathered1, aes(x = HAS_DIAGNOSIS, y=NUM_EVENTS, col=DIAGNOSIS)) +
    geom_col() +
    facet_grid(~PREVHYP) +
  labs(title = "Period 1: Events x Diagnosis, Diagnosis Type for Patients with and without Hypertension")
```



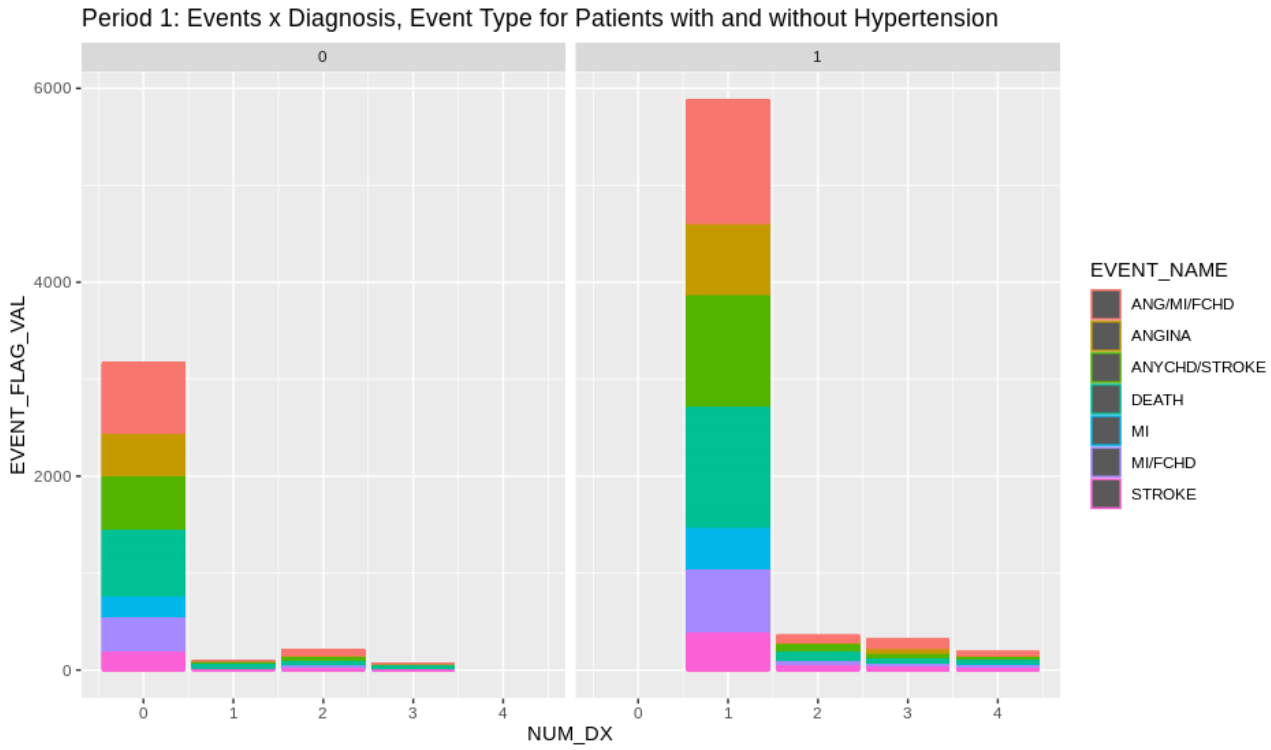
Patients with Hypertension had more events and more comorbidities (diagnoses), but it looks like the types of additional diagnoses were evenly distributed(and not mutually exclusive).

Let's look at a similar graph for period 1: Events x Diagnoses, without "Angina or MI" (since we already have Angina, and MI).

```
#Period 1 - Number & Type of Events by Number of Diagnoses; Facet Grid: PREVHYP
  ggplot(fr_gathered1, aes(x = NUM_DX, y=EVENT_FLAG_VAL, col=EVENT_NAME)) +
    geom_col() +
    facet_grid(~PREVHYP) +
  labs(title = "Period 1: Events x Diagnosis, Event Type for Patients with and without Hypertension")
```



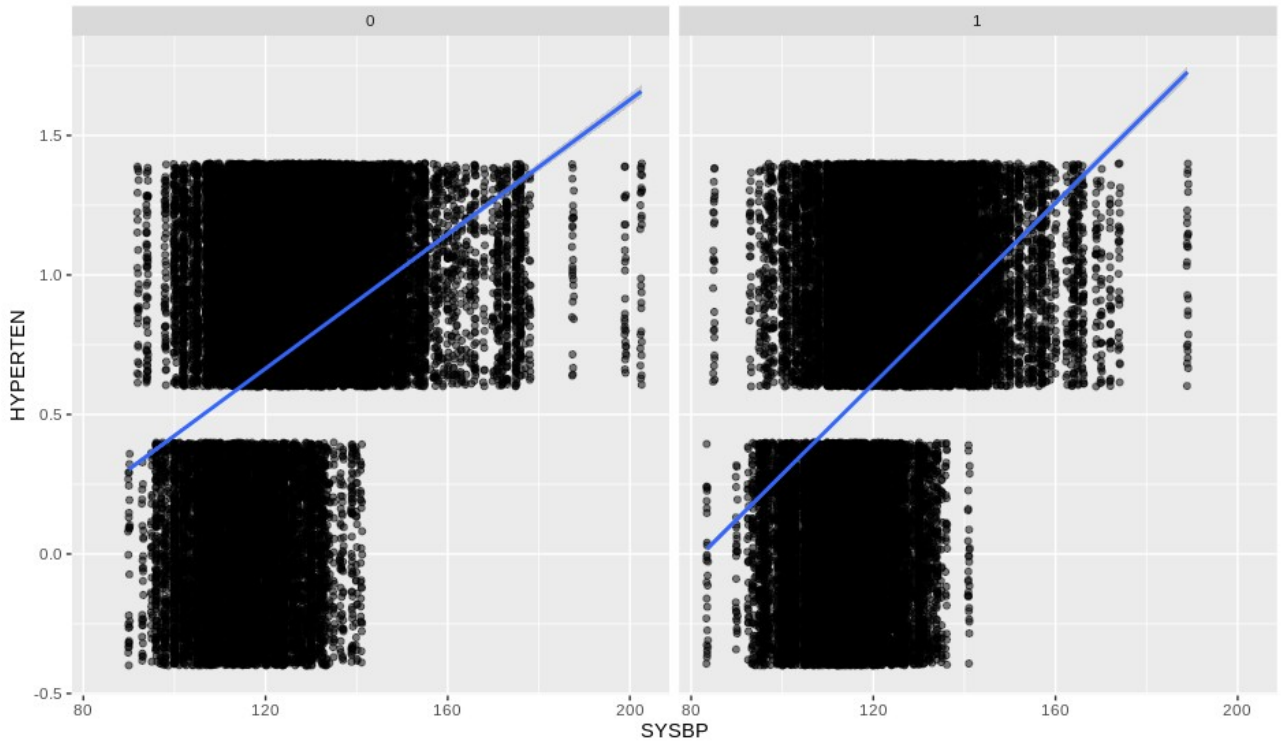
Let's add DEATH now, because as the number of diagnoses and events increases, it is likely death becomes a factor...



Measurable factors associated with Hypertension, by Period

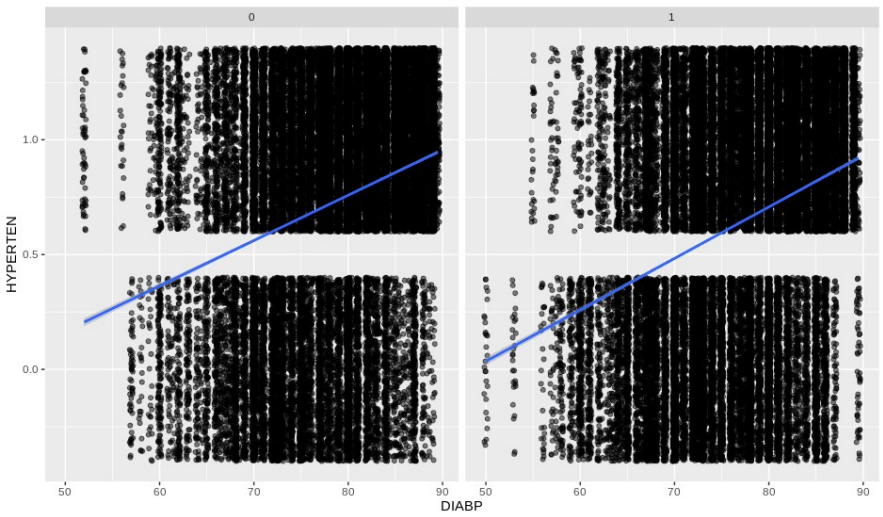
First, to look at Systolic and Diastolic BP, we need to go back to the original data set, before we recoded for Hypertension based on BP numbers:

```
ggplot(fr_gathered1, aes(x = SYSBP, y=HYPERTEN)) +
    geom_point(position="jitter", alpha=0.5) +
    geom_smooth(method = "lm") +
    facet_grid(~CURSMOKE)
    labs(title = "Period 1: Hypertension as Event x Systolic BP, Smoking Status")
```



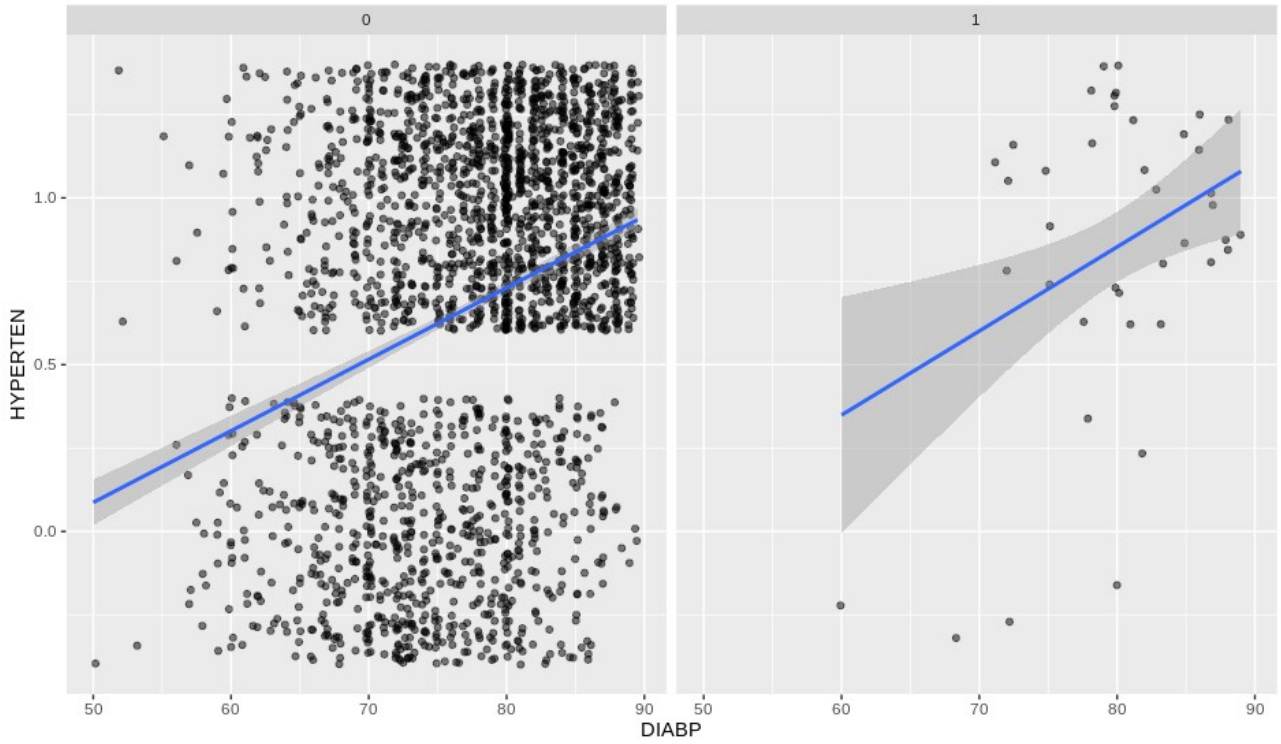
No surprises here. Systolic BP was strongly correlated with Hypertension, even more so for Smokers

What about diastolic pressure?



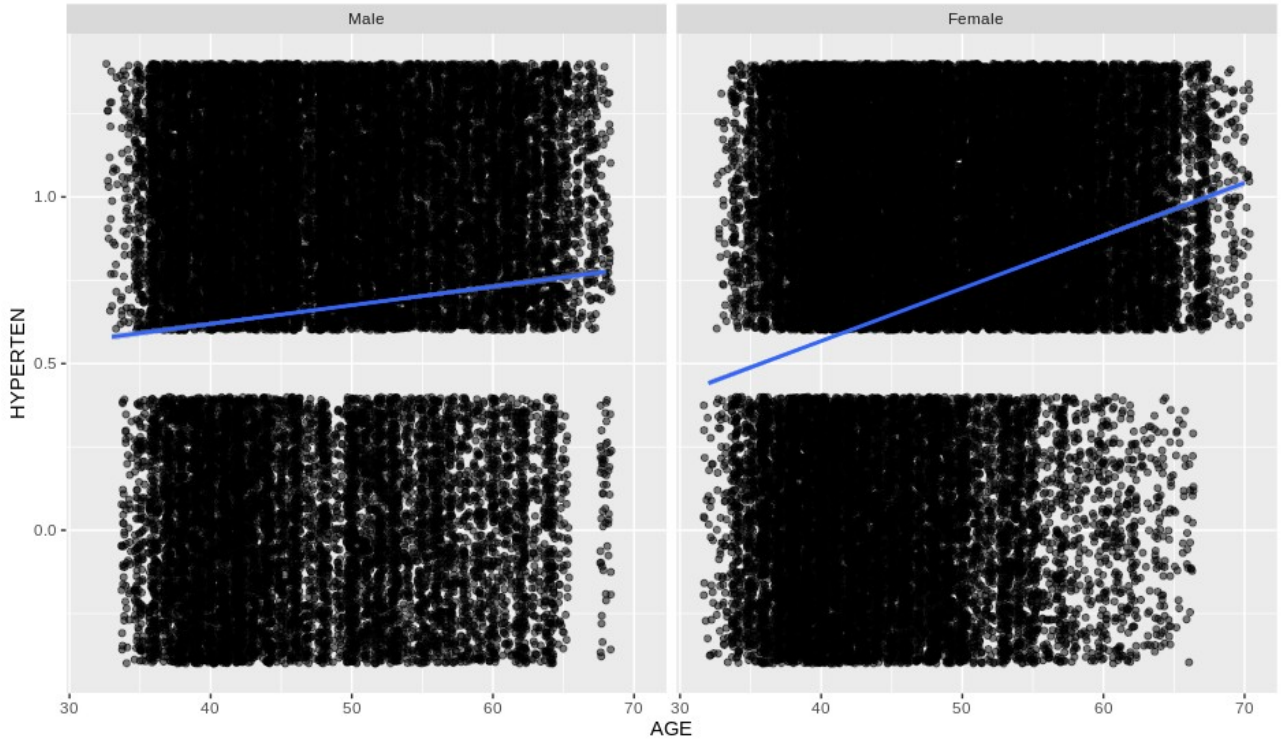
Looks like smoking didn't raise diastolic pressure in the same way.

But would Diabetes?



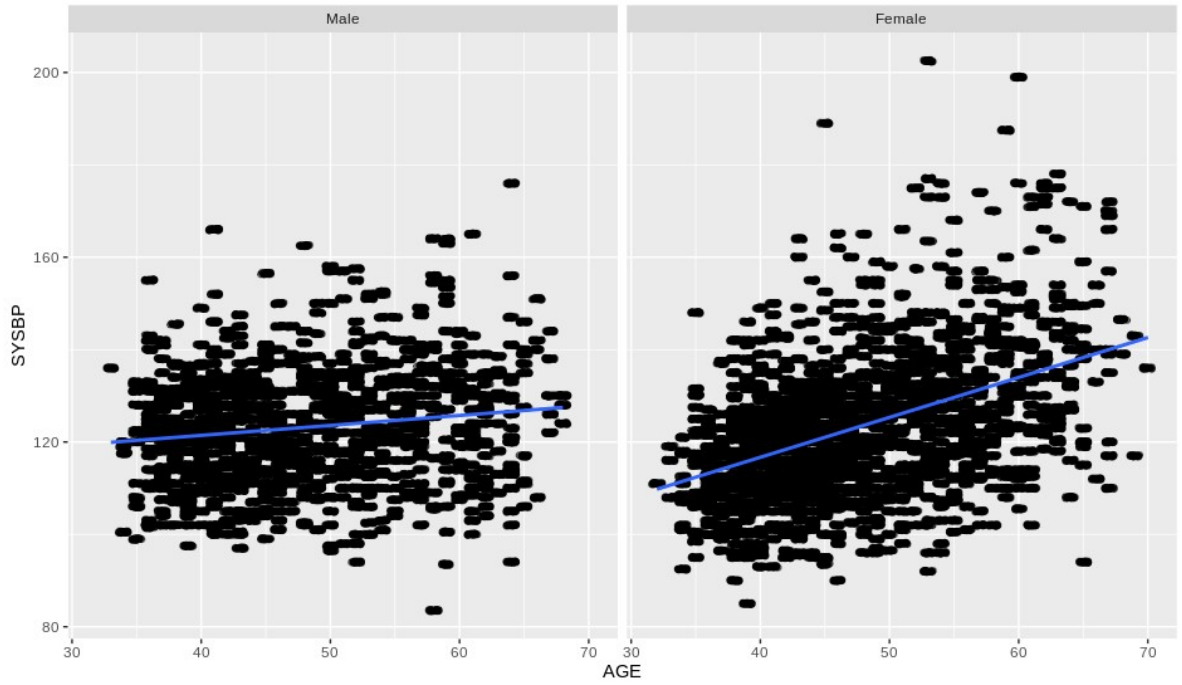
Hard to tell. It doesn't appear that way, but we also have a huge margin of error, and very few data points in the DIABETES positive group.

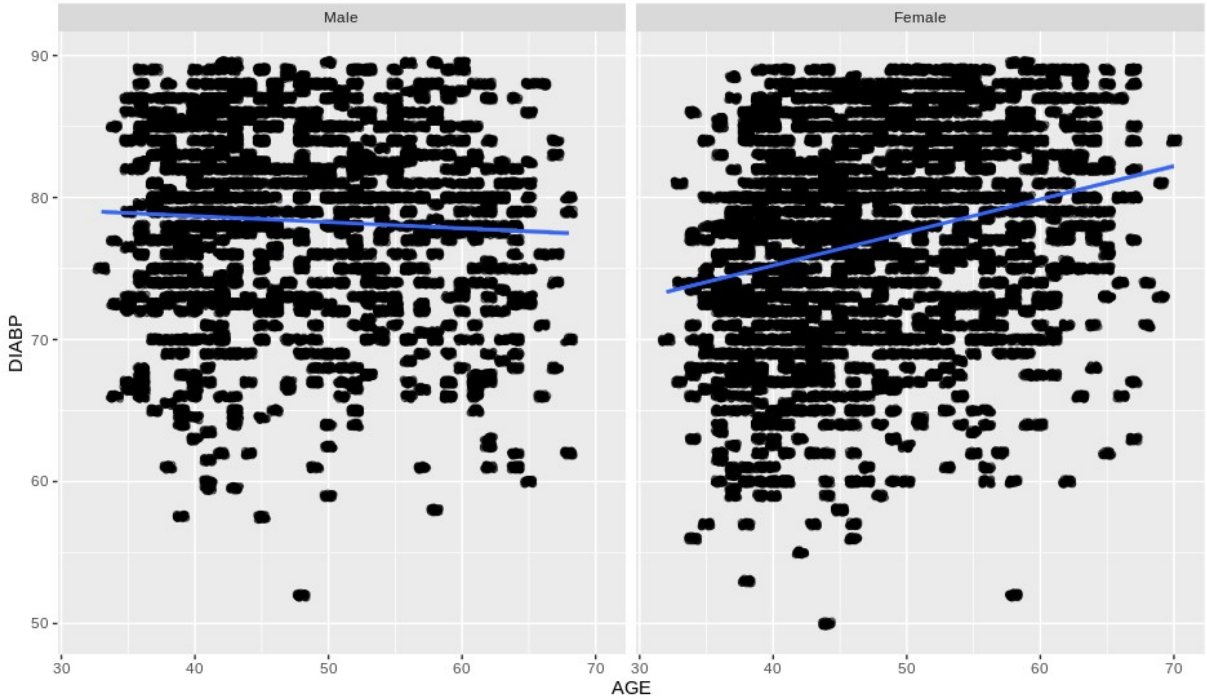
Differences between Males and Females by Age?



Here we can see how the protective effect of estrogen starts out protecting younger females, but as the female population aged, they not only caught up with males, they eventually surpassed them. Both Age and Sex appear to exert influence here.

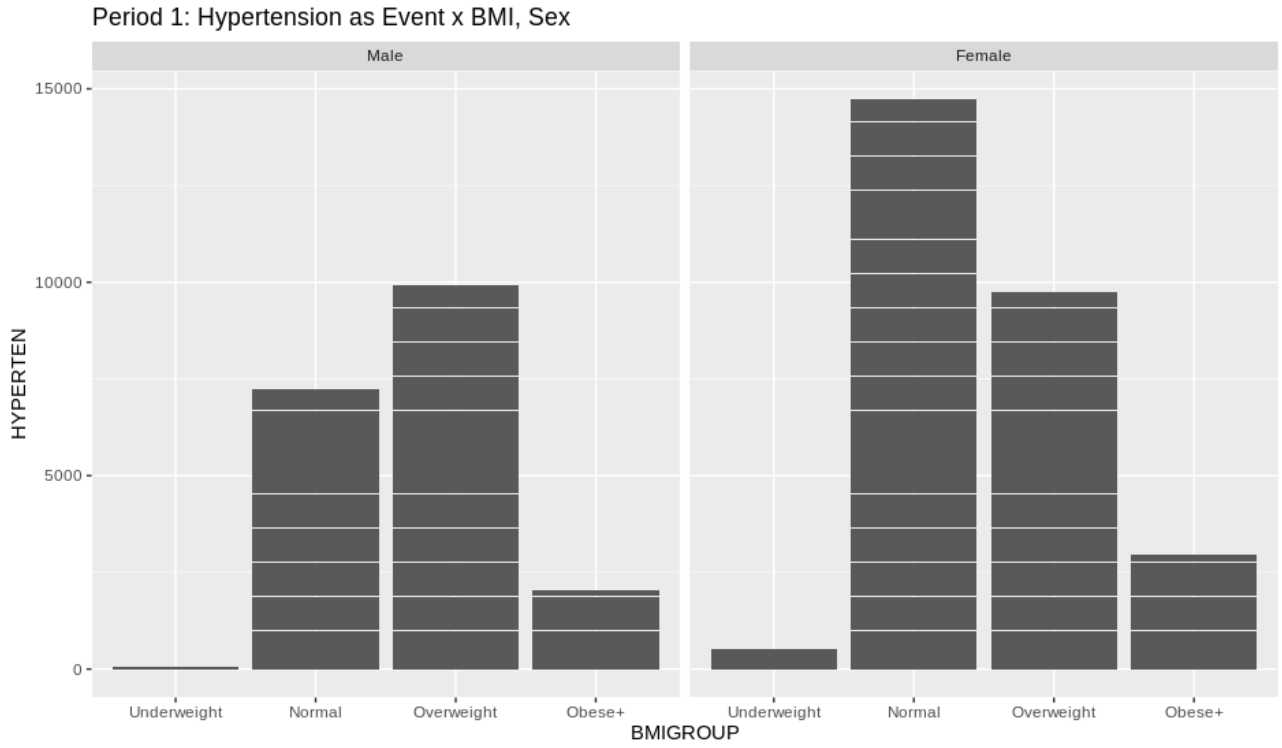
Let's look at Systolic and Diastolic pressure for the same effect:





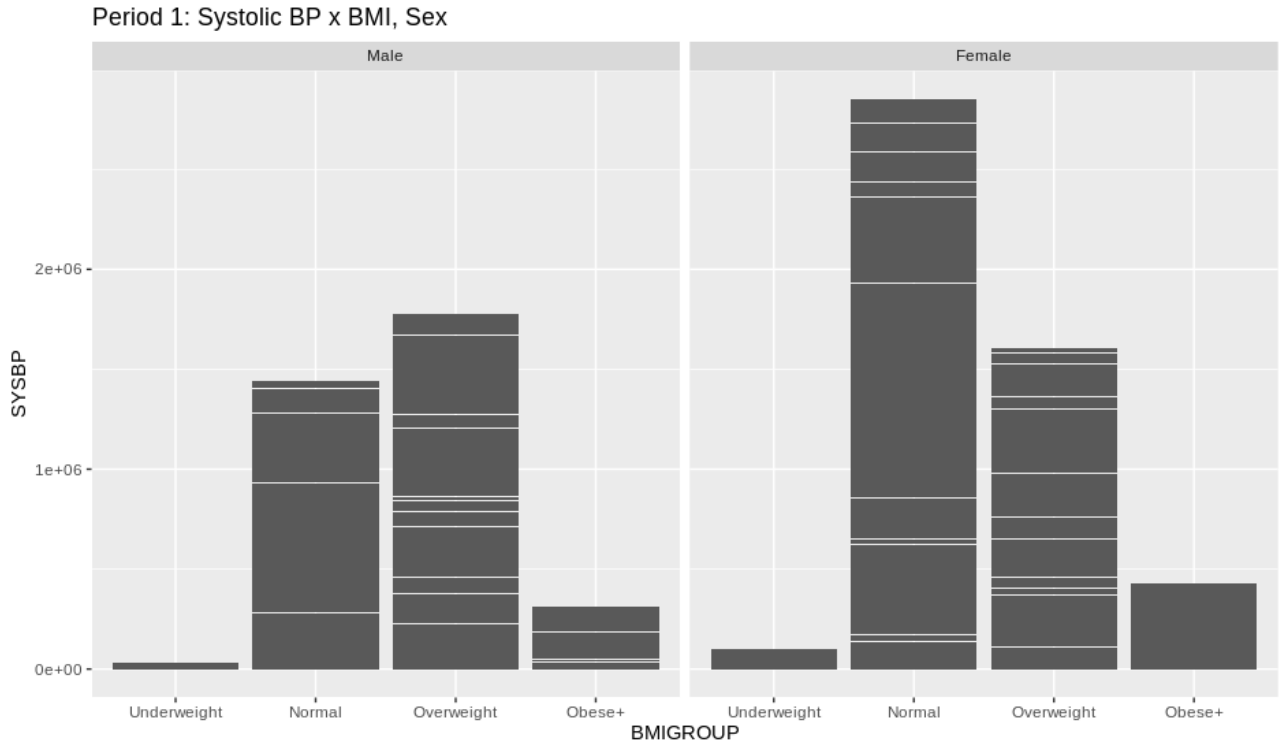
Again, we see the effect of Age on changes in Systolic and Diastolic pressure are greater for women. In the diastolic graph, we see men stay even and even decline in later years while women's pressure continued to increase!

Now, we need to look at the effects of our last two candidate explanatory variables: Glucose and BMI.

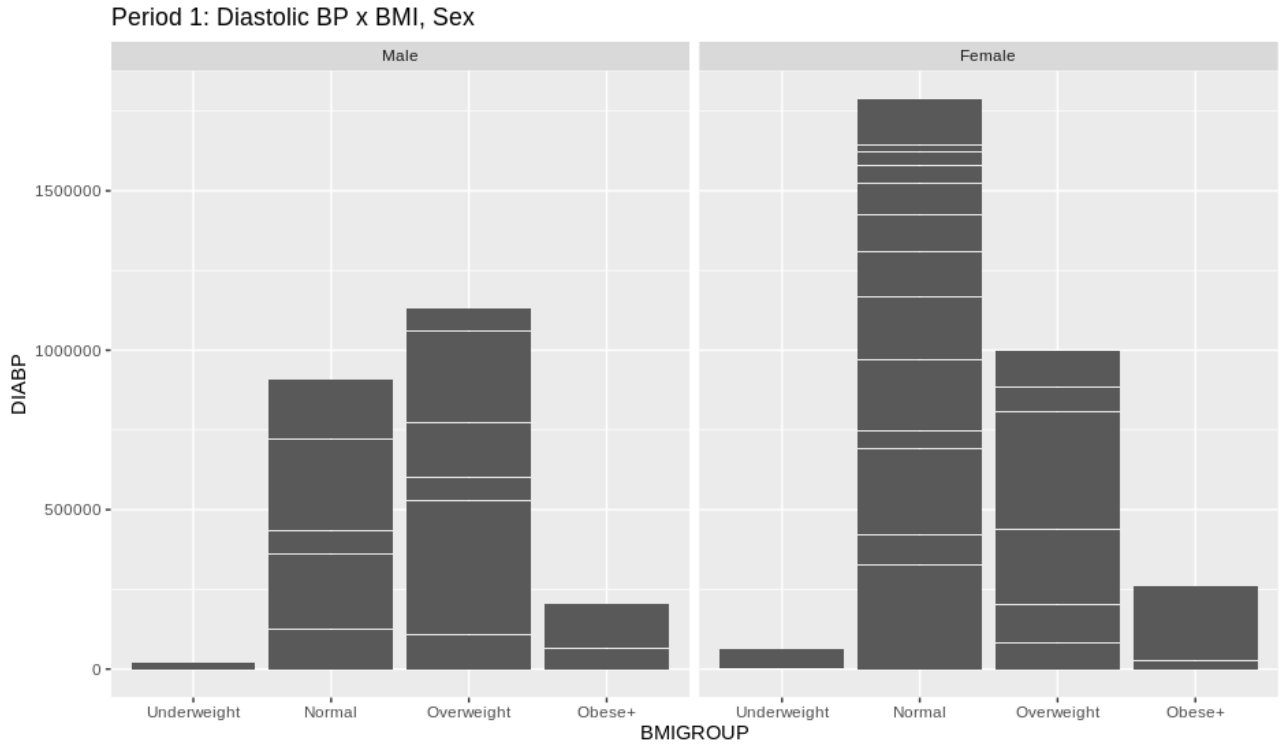


BMI appears to have an effect on hypertension as males go from normal weight to overweight, but for females, it appears this is not the reason for more hypertension, and age is more the factor.

Instead of looking at Event, let's look at Systolic and Diastolic pressure vs. BMI...



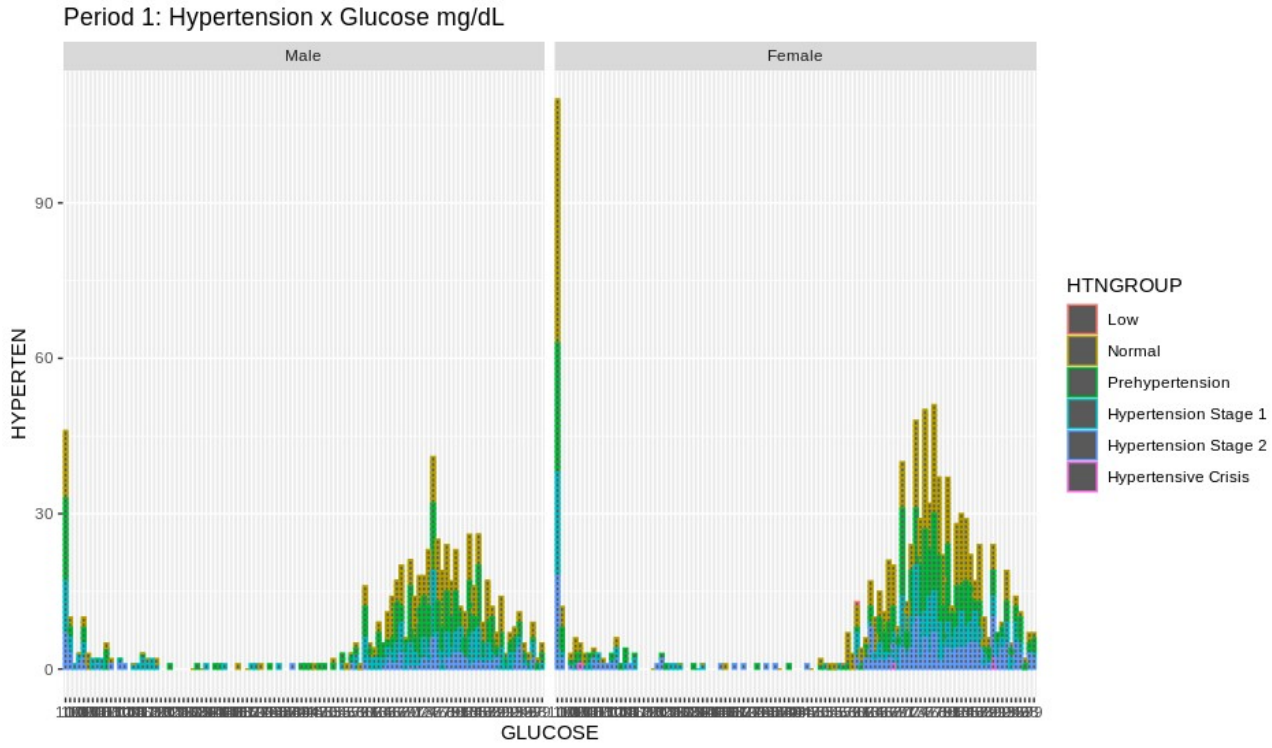
No change here.

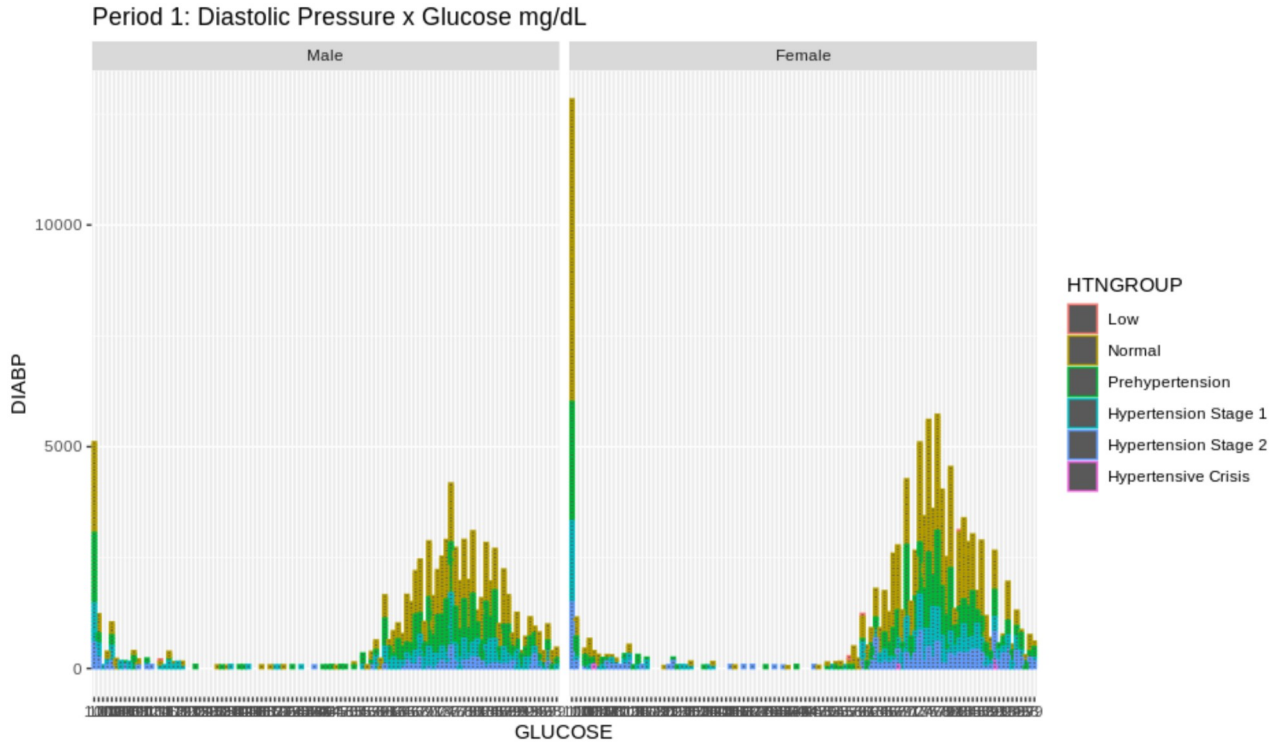


Nor here.

Now, Glucose. Attempts to remove persons with GLUCOSE > 200 or even 300 truncated 1000s of records. So, although the range of data was very large, I stayed with the entire data set. As you can see, the data here is not linear,

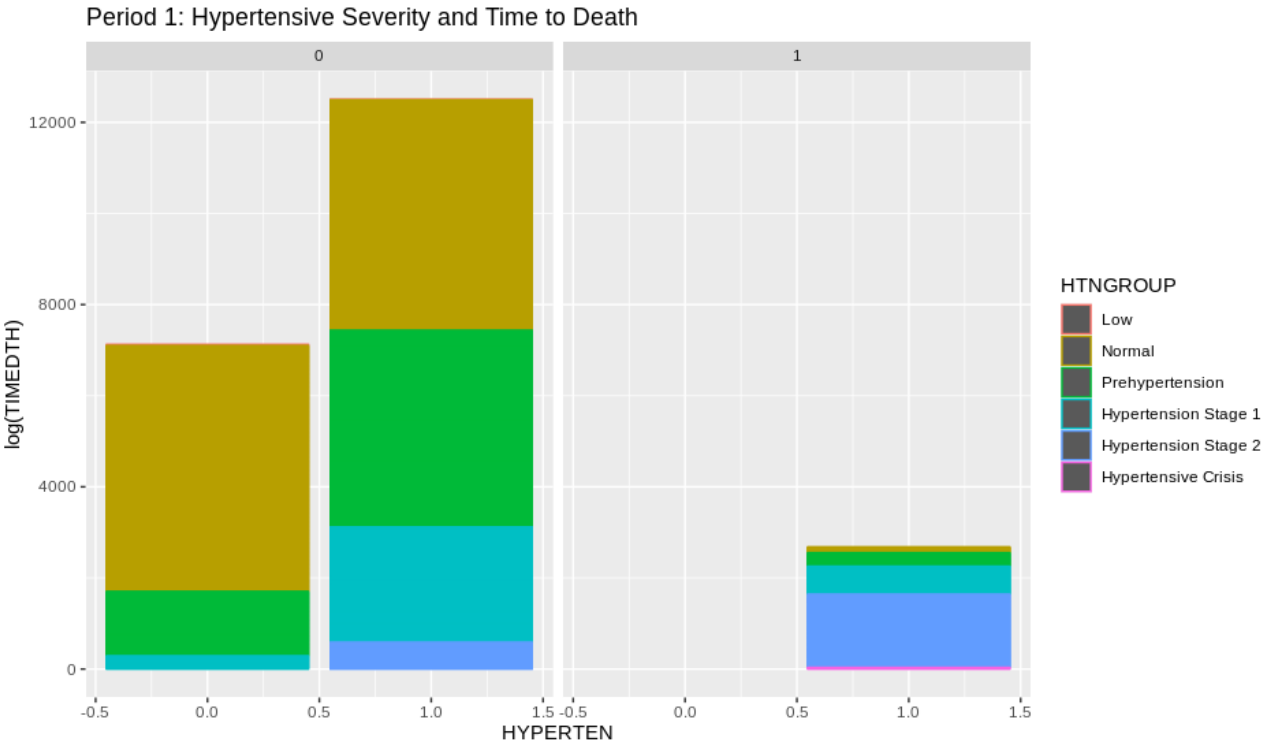
it is biphasic with a large # of hypertensive cases occurring with normal blood glucose, but a significant secondary spike occurring with blood glucose > 200.





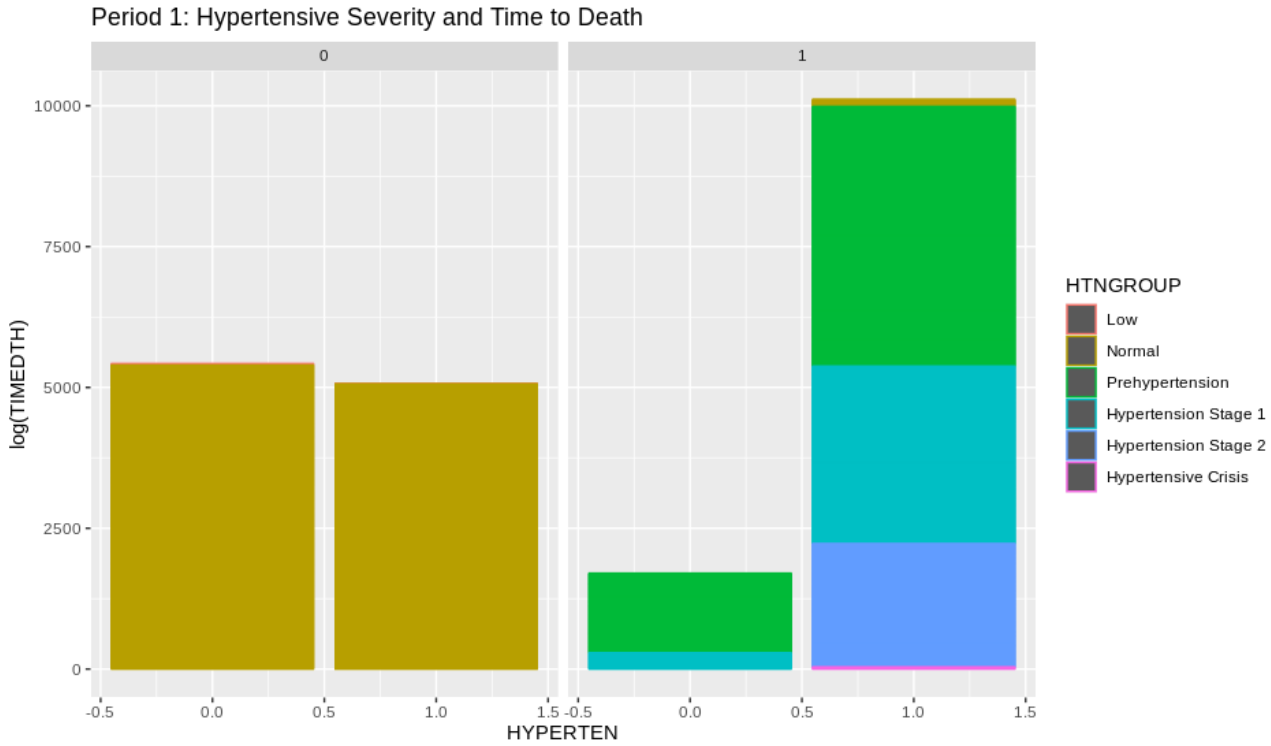
Last but not least, a peak at hypertensive severity and death

```
ggplot(fr_sum1, aes(x = HYPERTEN, y = log(TIMEDTH), col=HTNGROUP)) +
  geom_col() +
  facet_grid(~PREVHYP) +
  labs(title = "Period 1: Hypertensive Severity and Time to Death")
```



Those who had hypertension BEFORE the study began, had higher BPs overall, and the shortest time to Death. Those who began without hypertension and developed it (HYPERTEN = 1) lived longer, depending on how severe their HTN was. Notably, there were many in the undiagnosed group with Prehypertension or Stage 1 BP levels, who would not have been undiagnosed today.

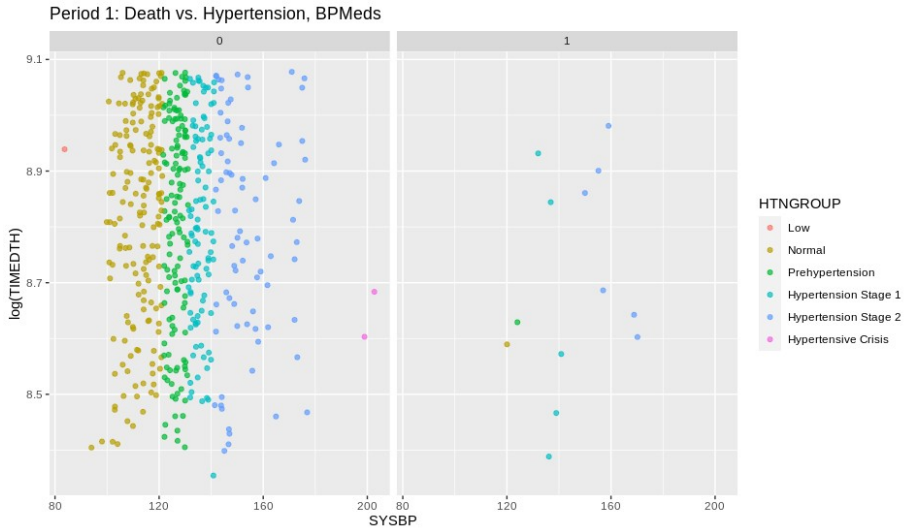
Recategorized based on BPs...



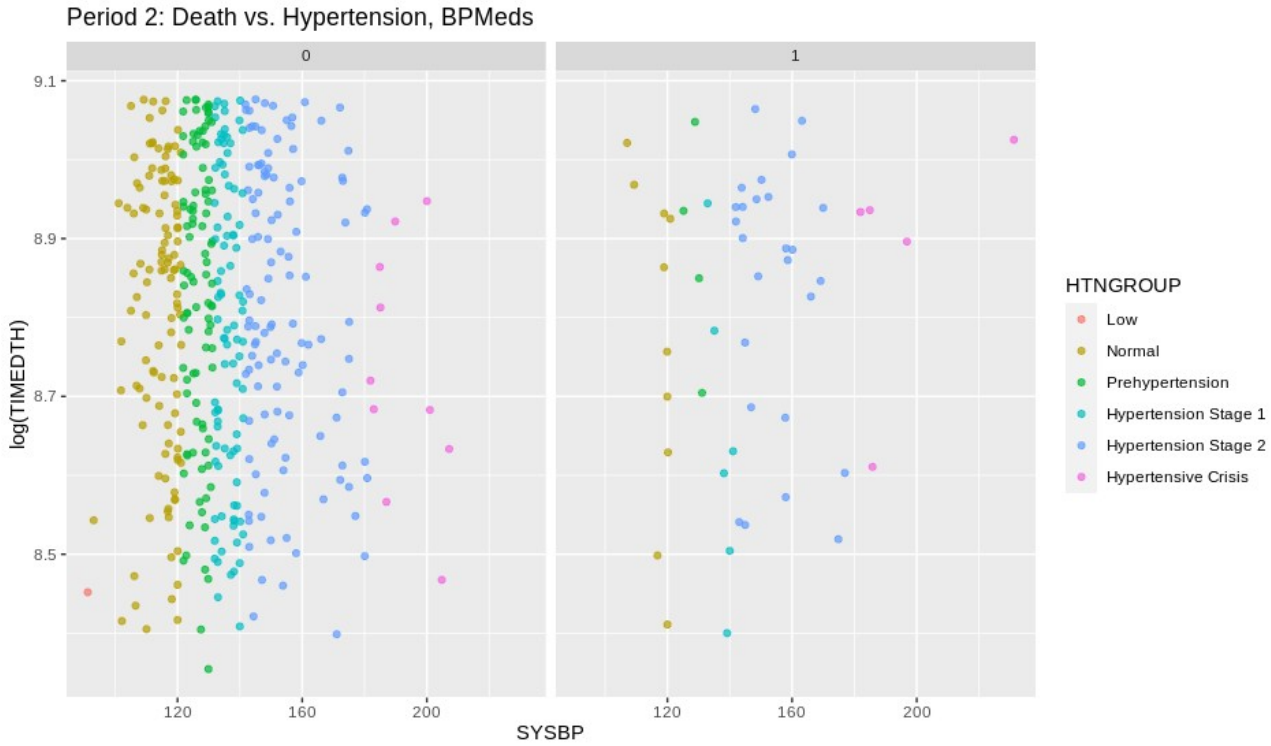
Makes a LOT more sense!

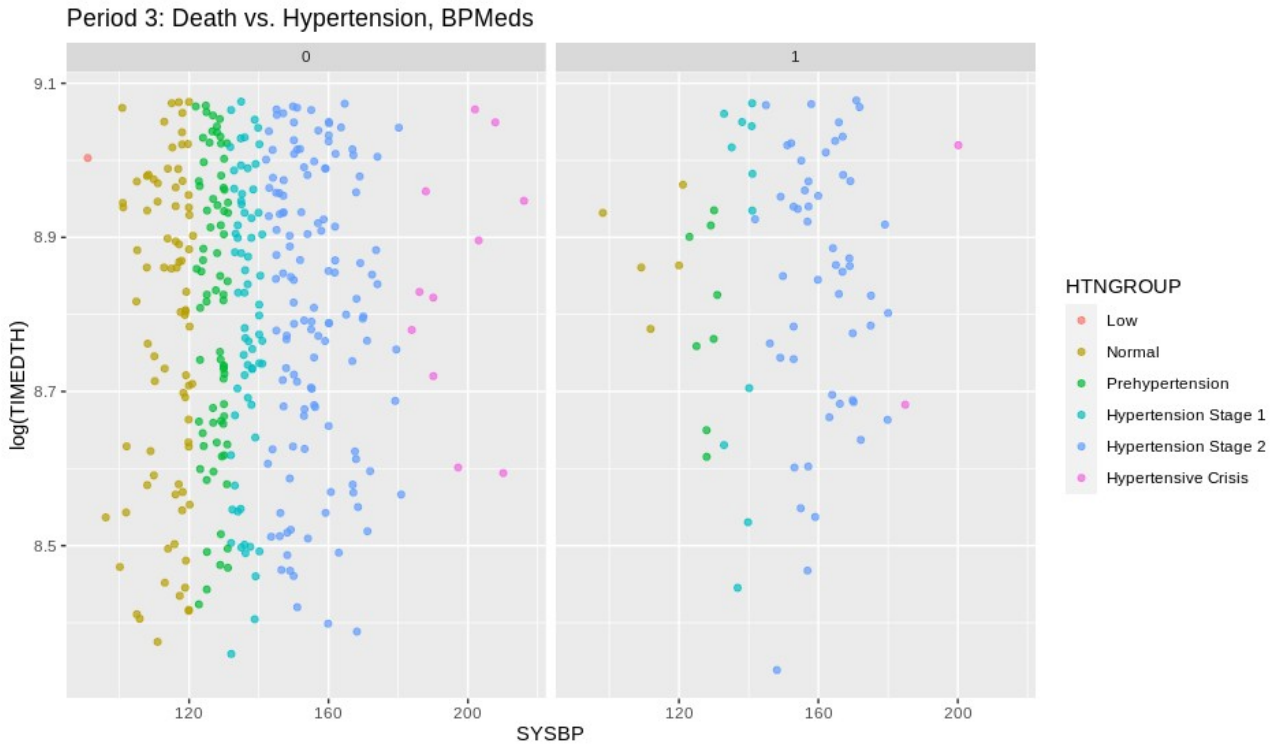
Everyone in the no HYPERTEN group has Normal BP (fancy that). And those who developed HYPERTEN had shorter time to death based on the severity of illness. We also need to keep in mind, many of the individuals marked NO DISEASE in this model may have been censored...

Let's filter out censored individuals, and then test for any mitigation from BP meds?



None of the periods had many individuals on BP meds, so it was impossible to dram a regression line for these graphs. However, there is definitely clustering at the top left for those with Normal and Prehypertension. Here are some comparisons with Period 2 and 3.





Variables identified as significant to the outcome of HYPERTENSION / TIME TO HYPERTENSION:

AGE, SEX, SYSBP (biggest effect), DIABP (colinear with SYSBP), GLUCOSE > 200, BMI > 30 (if Female)

CURSMOKE left out, because it has a direct effect on SYSBP, which accounts for its relationship with the outcome