Project Part I

What factors relate to high blood pressure?

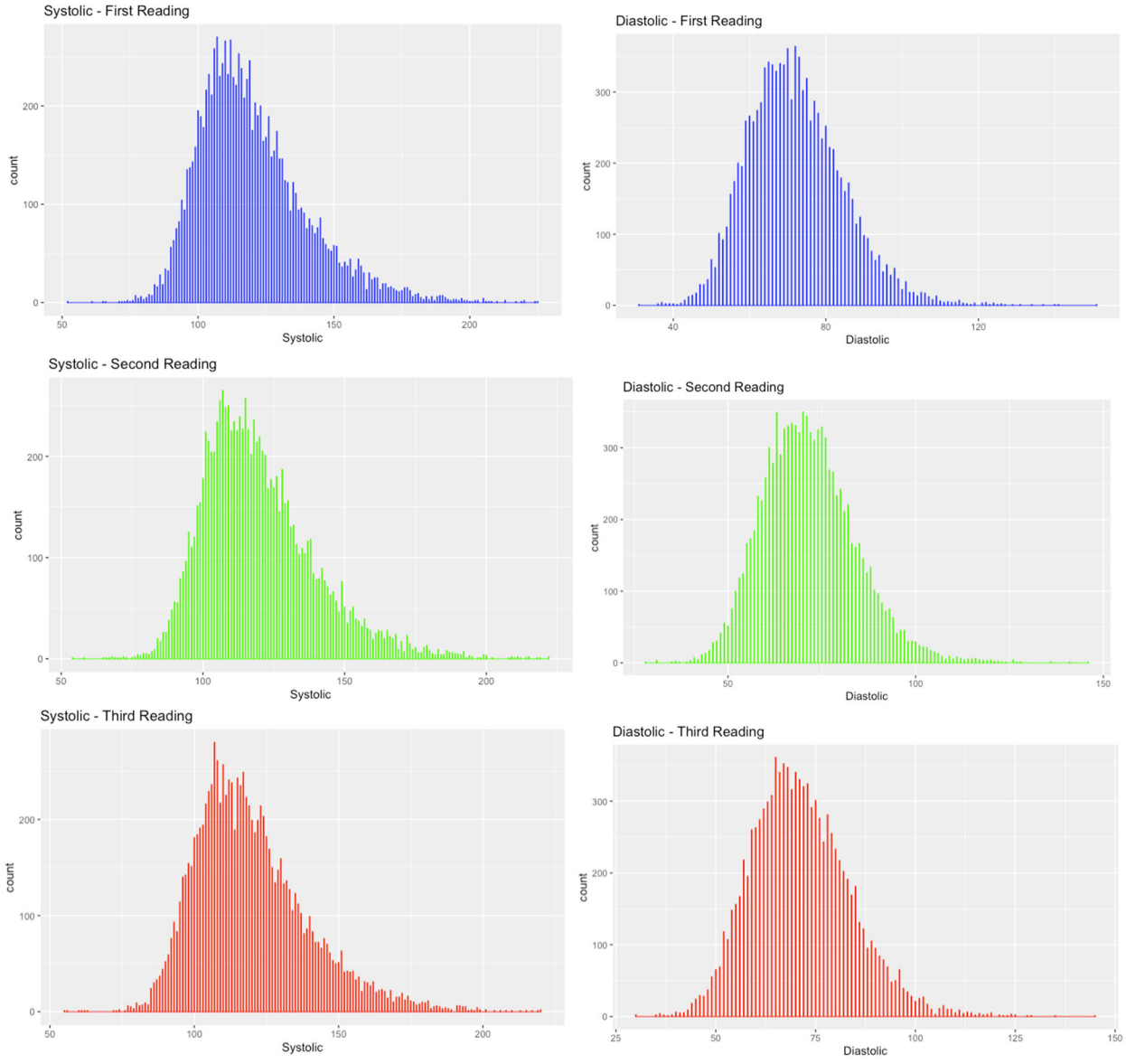
Purpose of Exploratory Analysis

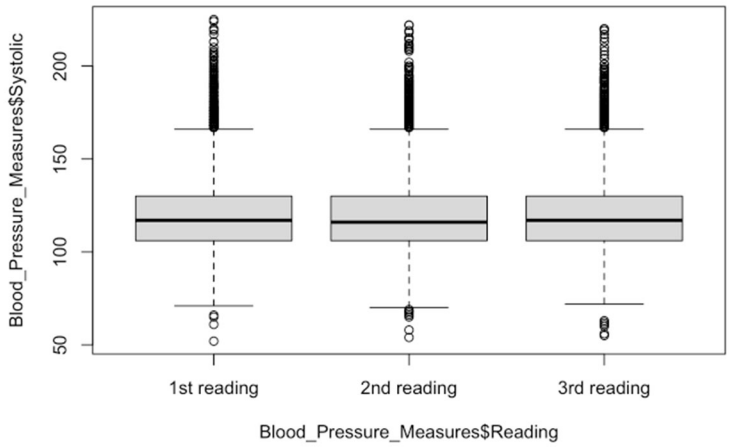
Medicine and health-related topics have always been interests of mine. In addition to that, several family members are currently experiencing hypertension. Hypertension can lead to complications such as stroke, heart attack, and dementia. In order to find out more about high blood pressure, I set the purpose of the analysis to explore factors that may relate to high systolic and diastolic blood pressures.

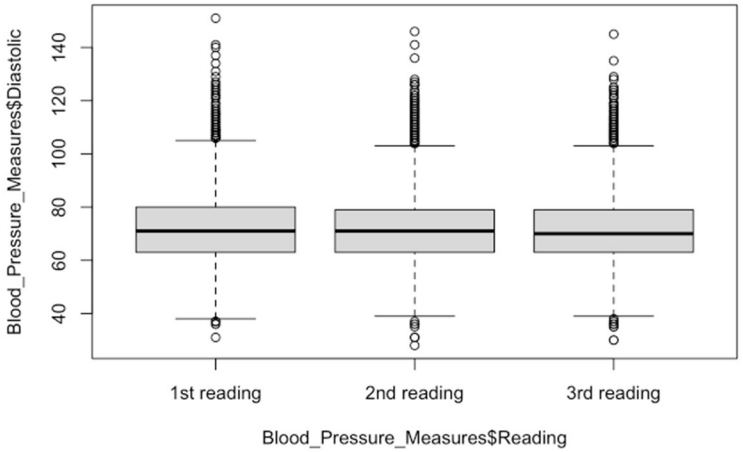
In order to do so, National Health and Nutrition Examination Survey 2017-March 2020 Pre-Pandemic data was downloaded from Centers for Disease Control and Prevention's website. Among the numerous datasets available, based on intuition, I have chosen eleven that seem to contain variables that may be related to blood pressure levels. The datasets are: Demographic Variables and Sample Weights, Dietary Supplement Use 30-Day - Total Dietary Supplements, Body Measures, Blood Pressure - Oscillometric Measurement, Blood Pressure & Cholesterol Questionnaire, Alcohol Use, Early Childhood, Health Insurance, Income, Sleep Disorders, and Occupation. Graphs such as histograms, box plots, scatter plots and summary statistics, as well as tables, are used to help identify potential relationships between the dependent variables, systolic and diastolic blood pressures, and the predictors.

Systolic Blood Pressure, Diastolic Blood Pressure

There are three sets of measurements in the Blood Pressure - Oscillometric Measurement dataset. It is necessary to check if they share a consistent distribution. Because if they do not, further investigation would be needed to decide the set of data to use to properly represent the general blood pressures of participants.







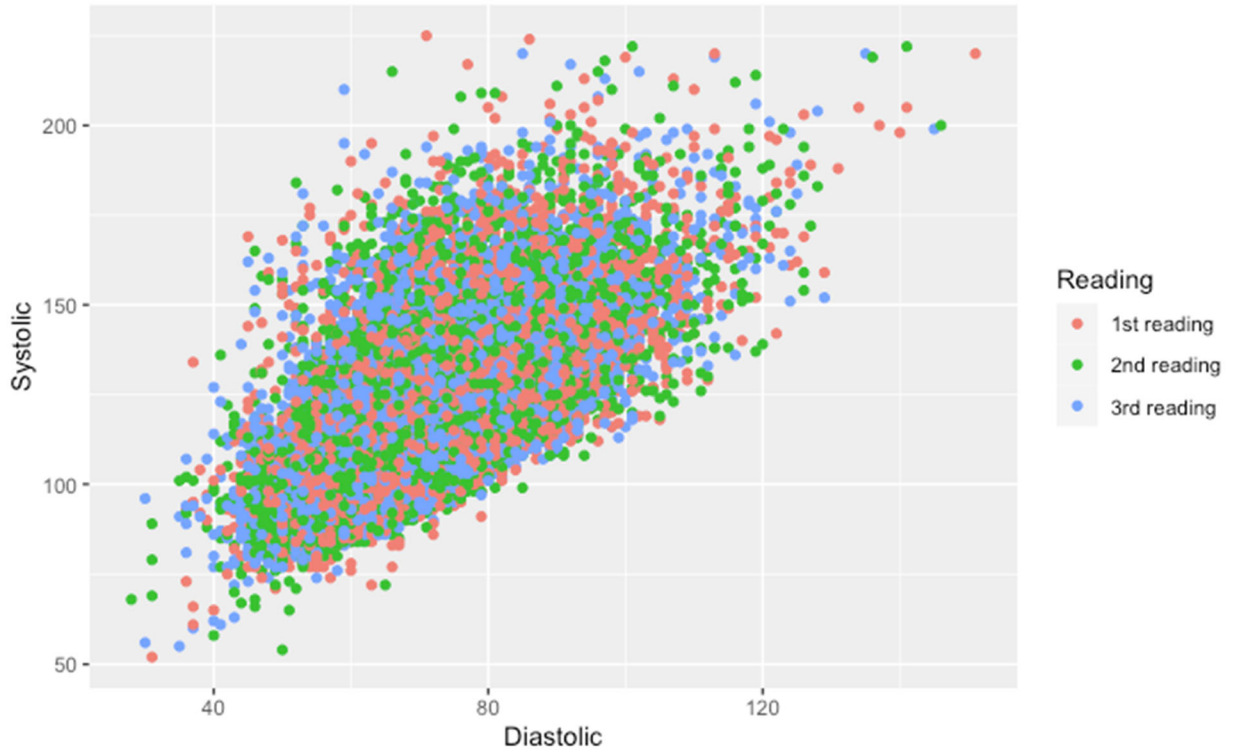

```
$systolic
$systolic$first
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
     52     106     117     120     130     225    1304

$systolic$second
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   54.0   106.0   116.0   119.7   130.0   222.0    1329

$systolic$third
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   55.0   106.0   117.0   119.7   130.0   220.0    1370
```

```
$diastolic
$diastolic$first
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  31.00   63.00   71.00   72.04   80.00  151.00    1304

$diastolic$second
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   28.0    63.0    71.0    71.5    79.0   146.0    1329

$diastolic$third
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  30.00   63.00   70.00   71.26   79.00  145.00    1370
```

Both of the graphs and summary statistics confirm that the three sets of blood pressure readings are fairly consistent: center of the distributions is around 120 mmHg for systolic blood pressure and is around 72 mmHg for diastolic; the spread is also almost identical through all three datasets - 50% of the data fall into the range of 106 - 130 mmHg for systolic and 63 - 80 mmHg for diastolic. This verification allows me to choose any set of readings as the dependent variable to work with the other variables.

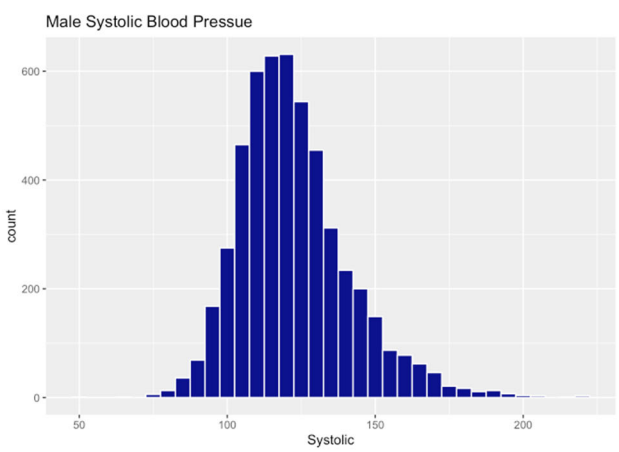
Systolic Blood Pressure vs. Diastolic Blood Pressure

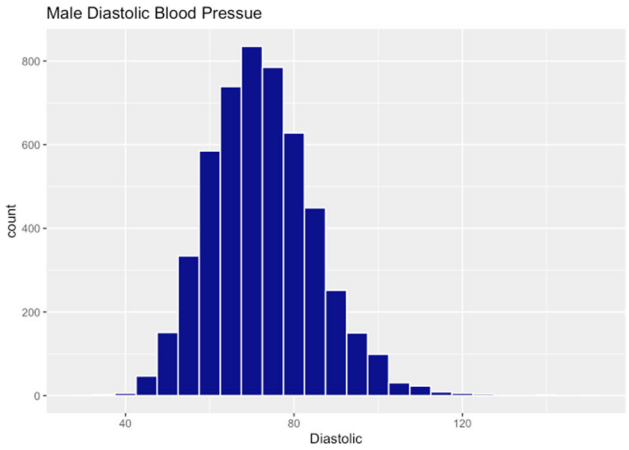
Common sense suggests that systolic blood pressure and diastolic blood pressure are highly and directly related; if one has high systolic blood pressure, then their diastolic reading should be high as well. The scatter plot below compares both types of blood pressure and confirms that the two types of blood pressures are directly related. Furthermore, all three sets of readings share this trend. The summary statistics here again show that all the record sets have the same means - 120 mmHg for systolic and 72 mmHg for diastolic, and the same variations - 20 mmHg for systolic and 12 mmHg for diastolic.

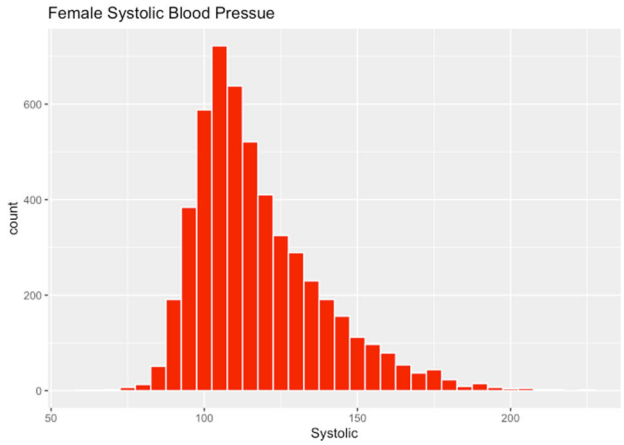


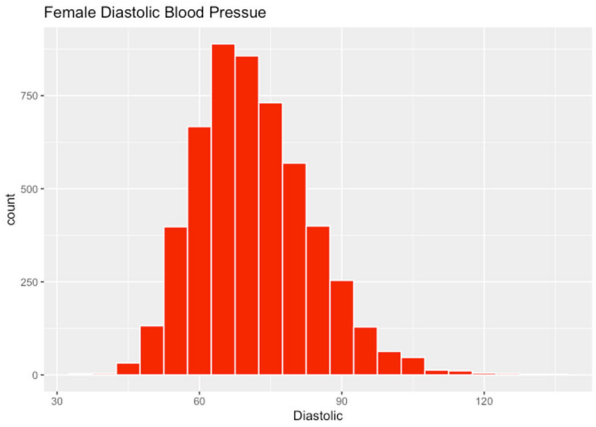
Reading <chr>	sys_mean <dbl>	sys_sd <dbl>	dia_mean <dbl>	dia_sd <dbl>
1st reading	120.0046	19.95013	72.03574	12.41688
2nd reading	119.7323	19.78850	71.50382	12.43684
3rd reading	119.6819	19.64162	71.25880	12.40516

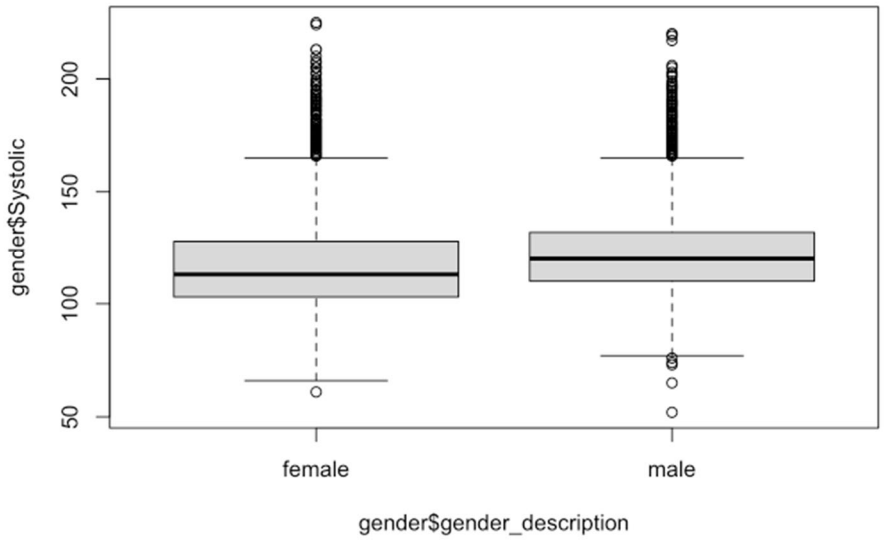
Gender vs. Blood Pressures

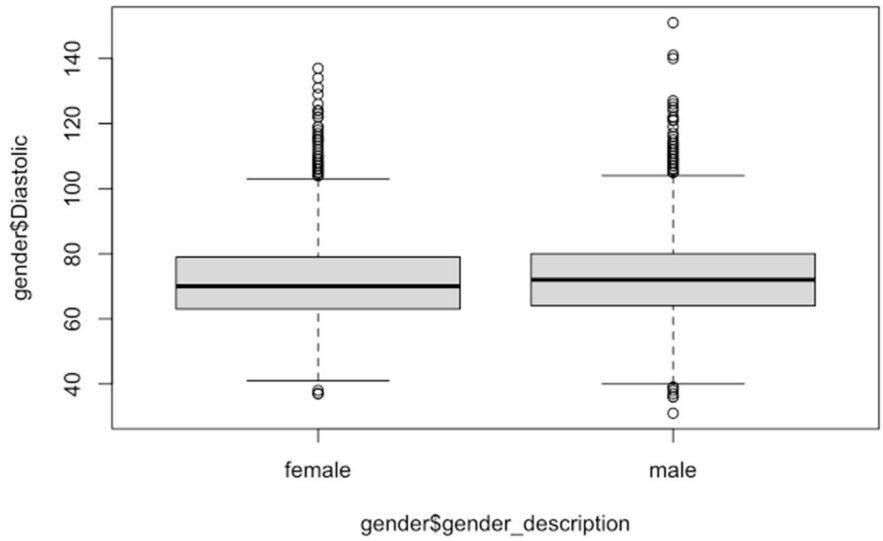












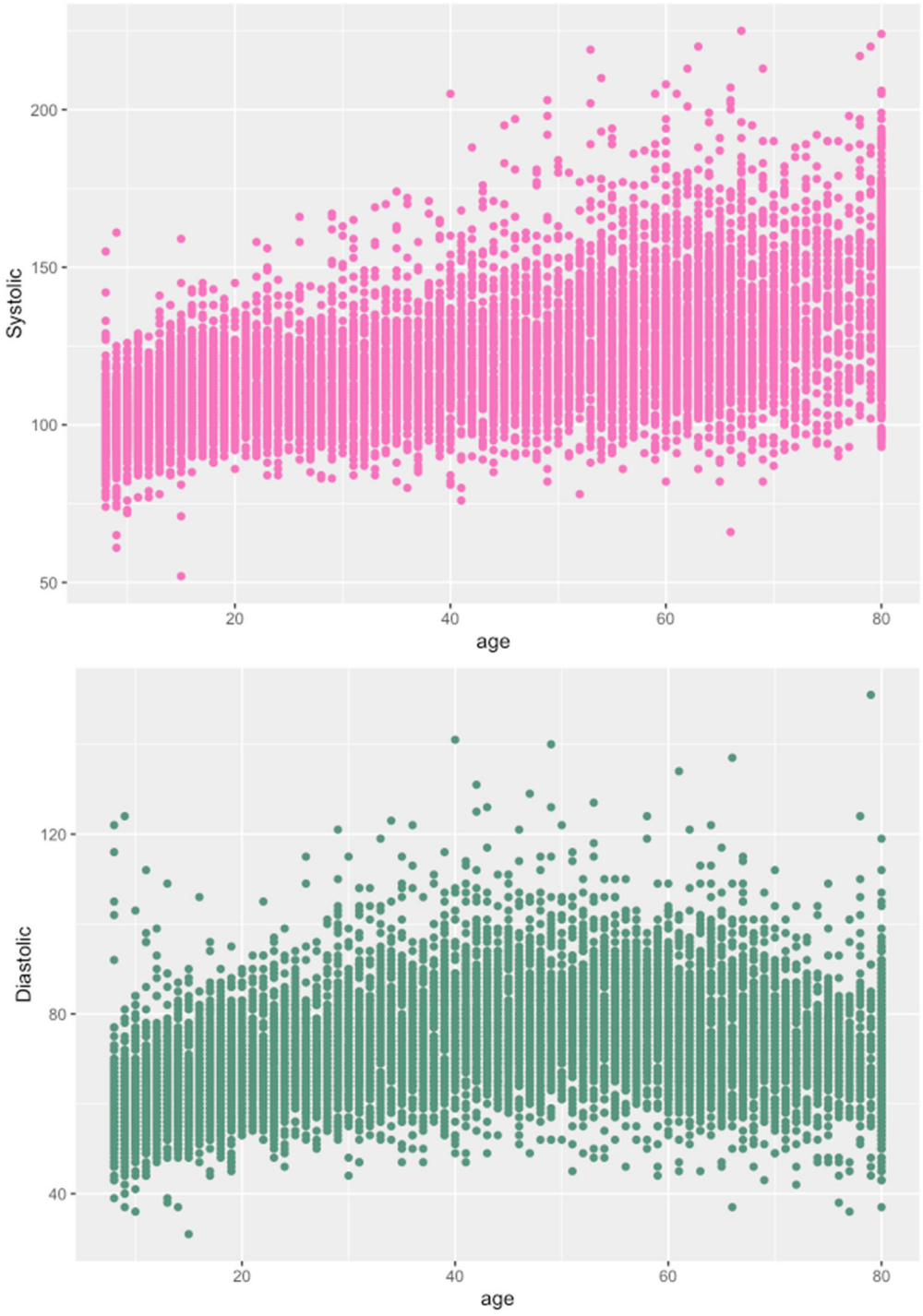

```
$systolic
$systolic$name
[1] "male"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   52.0   110.0   120.0   122.5   132.0   220.0     584

$systolic$name
[1] "female"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   61.0   103.0   113.0   117.6   128.0   225.0     720
```

```
$diastolic
$diastolic$name
[1] "male"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  31.00   64.00   72.00   72.48   80.00  151.00     584

$diastolic$name
[1] "female"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   37.0    63.0    70.0    71.6    79.0   137.0     720
```

Here, the gender variable is explored. In general, do men have a higher blood pressure level than women? The histograms of blood pressures distinguished by genders show that the female-systolic histogram is skewed to the right. This means the majority of the readings fall in the lower blood pressure range. Compared to this distribution, the male-systolic histogram is less skewed and seems to have a higher number for the center. The shapes of the diastolic histograms look similar for both genders, but the centers are different - male's seems to have a higher mean again than that of the female's.

The box plots compare both genders' distributions side by side and clearly demonstrate that females' means are smaller than that of male's and females' systolic blood pressures have a wider variance than male's. The summary statistics confirm the observations of the graphs: female's systolic median is 113 mmHg, much smaller than male's 120 mm Hg; 50% of female's data points fall into the 103-128 mmHg range, 3 mmHg wider than male's 110-132 mmHg range; female's diastolic mean is 71.6 mmHg, which is again smaller than male's 72.48 mmHg, although the difference may not be significant; the center 50% of the female diastolic data points are in the 63-79 mmHg range, which has the same spread compared to male's 64-80 mmHg range.

Age vs. Blood Pressures



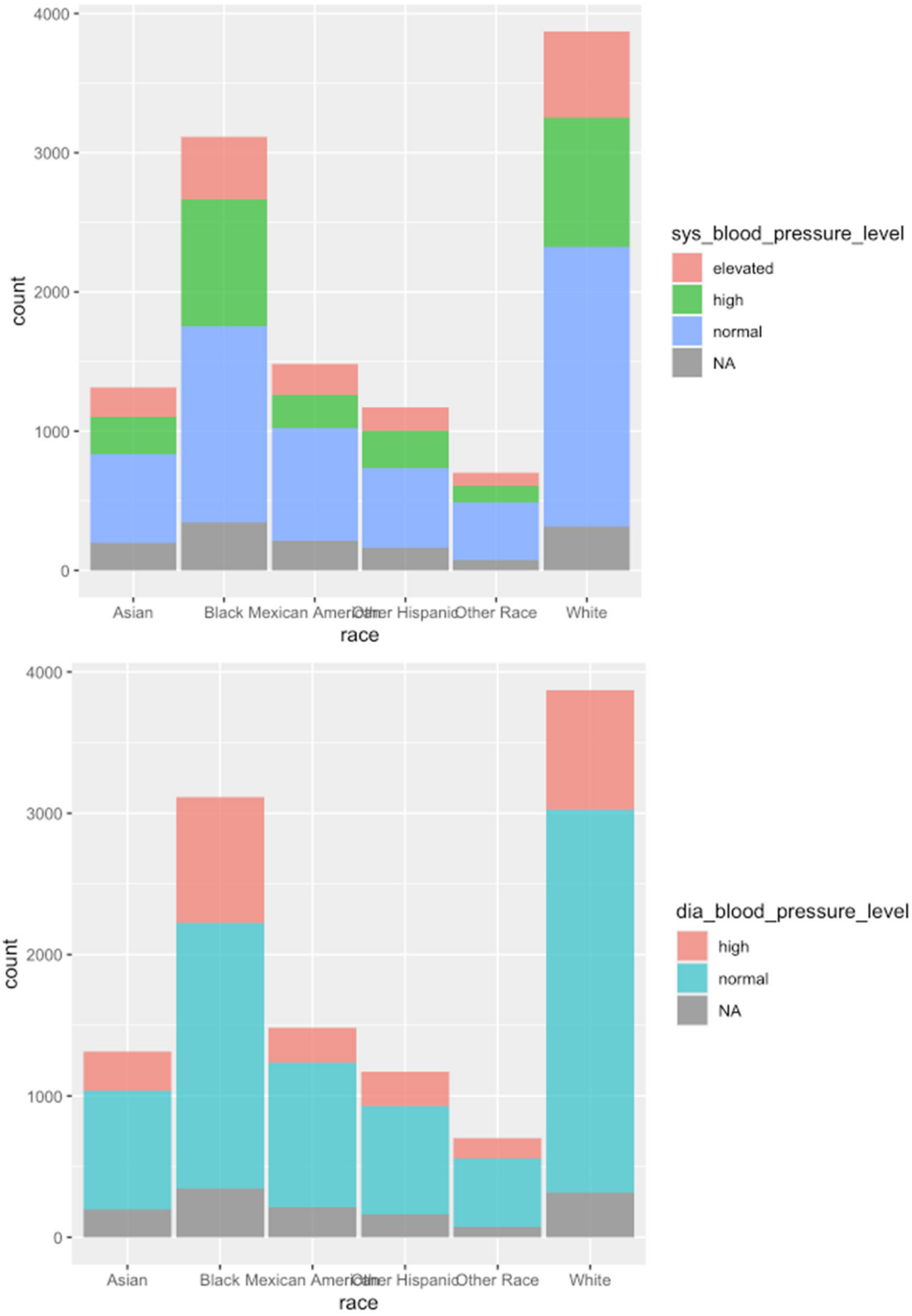
The next variable to explore is age. Do older people have higher blood pressure? The systolic scatter plot shows that indeed, there is a clear direct relationship between systolic blood pressure and age. However, the diastolic scatter plot shows the same trend only up to age 50. After 50, the relationship reverses; the two variables become inversely related. Blood pressure summary statistics by age

groups further confirm these trends. The means from the youngest age group to the oldest are 100.8, 108.1, 115, 127.4, and 136.1 mmHg for systolic, and 60.11, 64.36, 72.97, 78.3, and 72.87 mmHg for diastolic. Also, it can be observed that younger people's systolic blood pressure interquartile ranges are narrower than that of older people. With diastolic blood pressure, children and teens seem to have the same, narrower interquartile ranges, and those who are older have the same, wider spread. The summary statistics below align with such observation.

```
$systolic
$systolic$name
[1] "infants"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.


$systolic$name
[1] "children"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   61.0    94.0   101.0   100.8   107.0   161.0     196

$systolic$name
[1] "teens"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   52.0   101.0   108.0   108.1   114.0   159.0     167

$systolic$name
[1] "young adults"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
     76     105     114     115     123     205     413

$systolic$name
[1] "middle age"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   78.0   114.0   125.0   127.4   138.0   220.0     263

$systolic$name
[1] "65+"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   66.0   121.0   134.0   136.1   150.0   225.0     265
```

```
$diastolic
$diastolic$name
[1] "infants"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.


$diastolic$name
[1] "children"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  36.00   55.00   59.00   60.11   65.00  124.00     196

$diastolic$name
[1] "teens"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  31.00   59.00   64.00   64.36   69.00  109.00     167

$diastolic$name
[1] "young adults"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  44.00   65.00   72.00   72.97   80.00  141.00     413

$diastolic$name
[1] "middle age"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   44.0    70.0    78.0    78.3    86.0   140.0     263

$diastolic$name
[1] "65+"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  36.00   65.00   72.00   72.87   80.00  151.00     265
```

Race vs. Blood Pressures

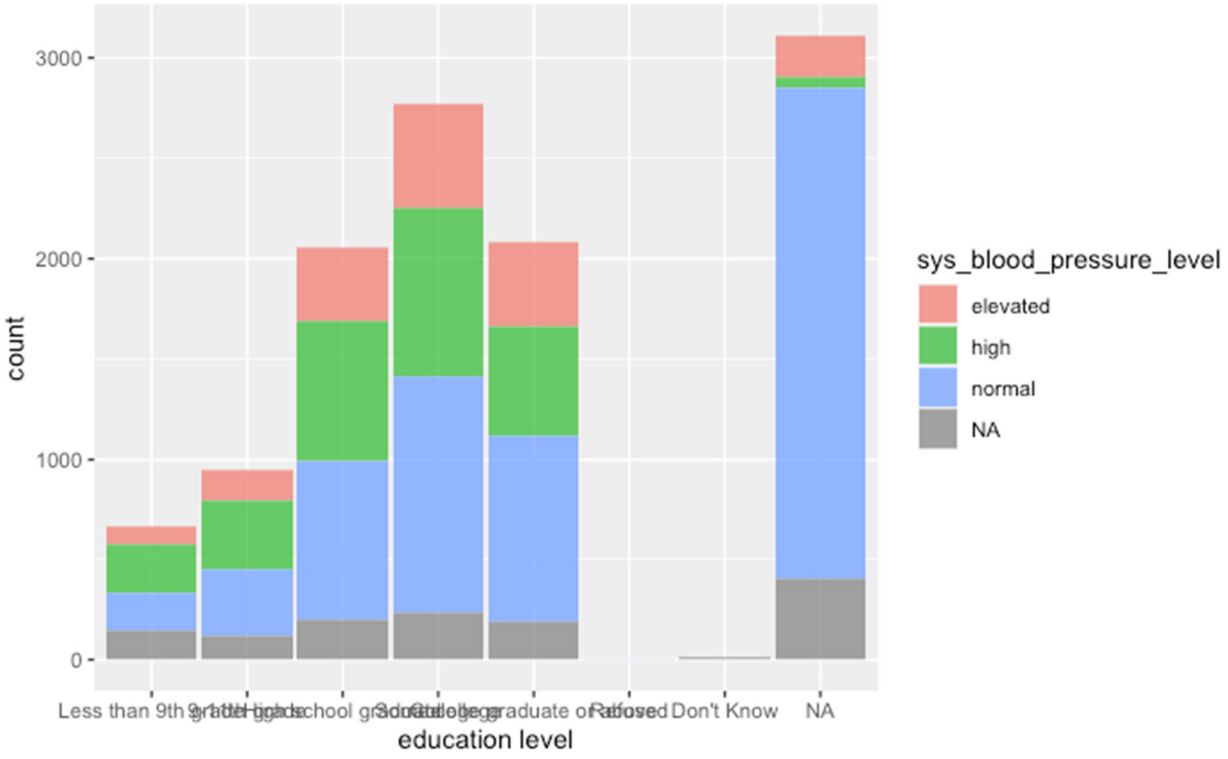


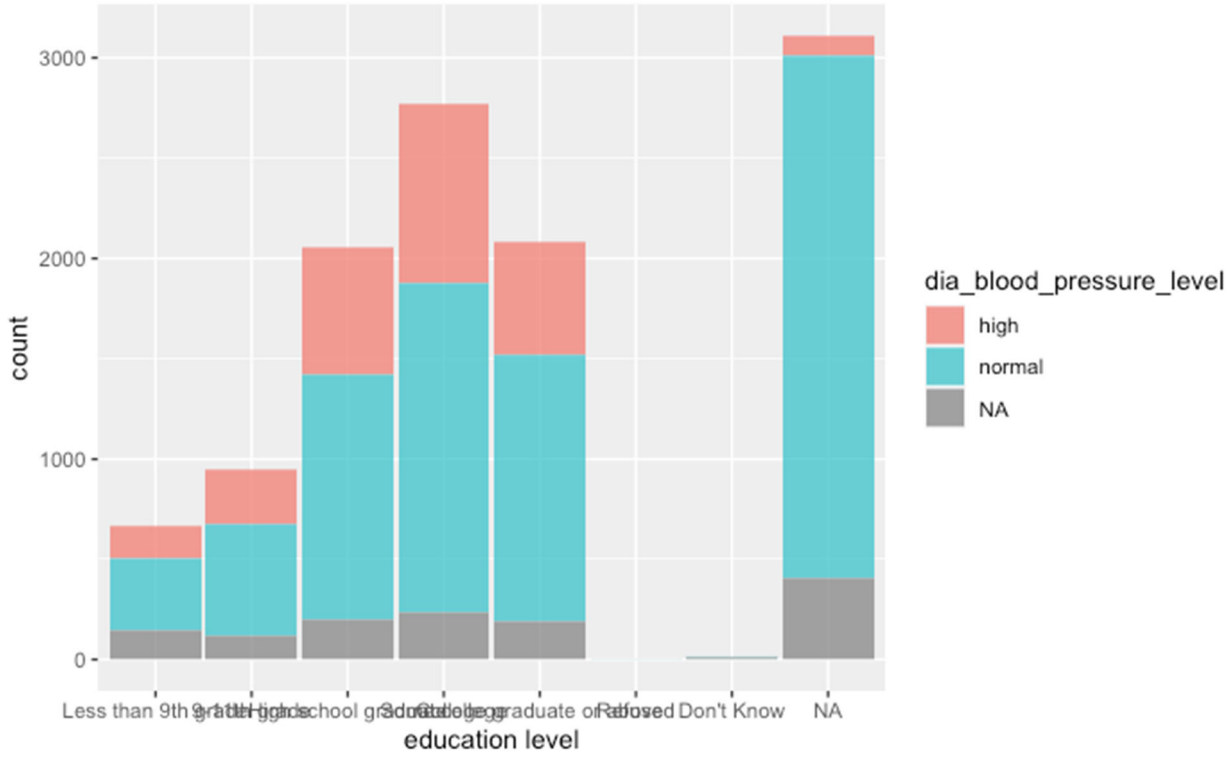
	elevated	high	normal	<NA>	Sum
Asian	212	267	640	196	1315
Black	449	910	1408	346	3113
Mexican American	222	238	812	210	1482
Other Hispanic	168	269	575	161	1173
Other Race	94	120	414	73	701
White	620	929	2005	318	3872
<NA>	0	0	0	0	0
Sum	1765	2733	5854	1304	11656

	high	normal	<NA>	Sum
Asian	276	843	196	1315
Black	891	1876	346	3113
Mexican American	247	1025	210	1482
Other Hispanic	242	770	161	1173
Other Race	145	483	73	701
White	848	2706	318	3872
<NA>	0	0	0	0
Sum	2649	7703	1304	11656

The next variable to explore is race. Does a particular race have higher blood pressure than the rest? Bar charts and count tables are created to compare proportions. In the systolic blood pressure and race bar chart, black people seem to have the biggest proportion - 43.66% based on the count table, in high and elevated blood pressure levels combined, the next highest in those levels is white people - 40%. Both of these percentages are higher than 38.59%, the proportion that is in the two levels in the whole population. While other race has the smallest proportion in high and elevated levels - 30.53%, and the largest proportion in normal - 59.06%. For diastolic blood pressure, black people again have the largest proportion in the high level - 28.62%, which is significantly above the 22.73% population proportion. The second largest is white people - 21.9%. Mexican Americans have the smallest proportion in this level - 16.67%, and the highest percentage of normal blood pressure - 69.16%.

Education vs. Blood Pressures



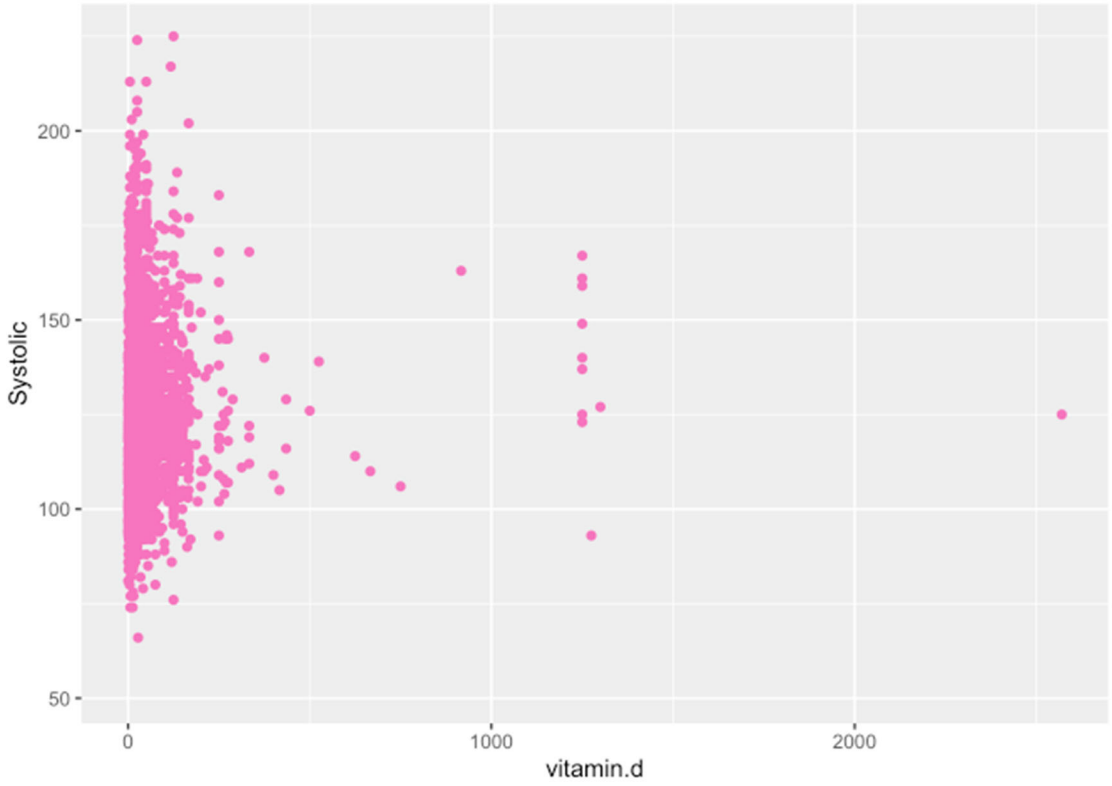


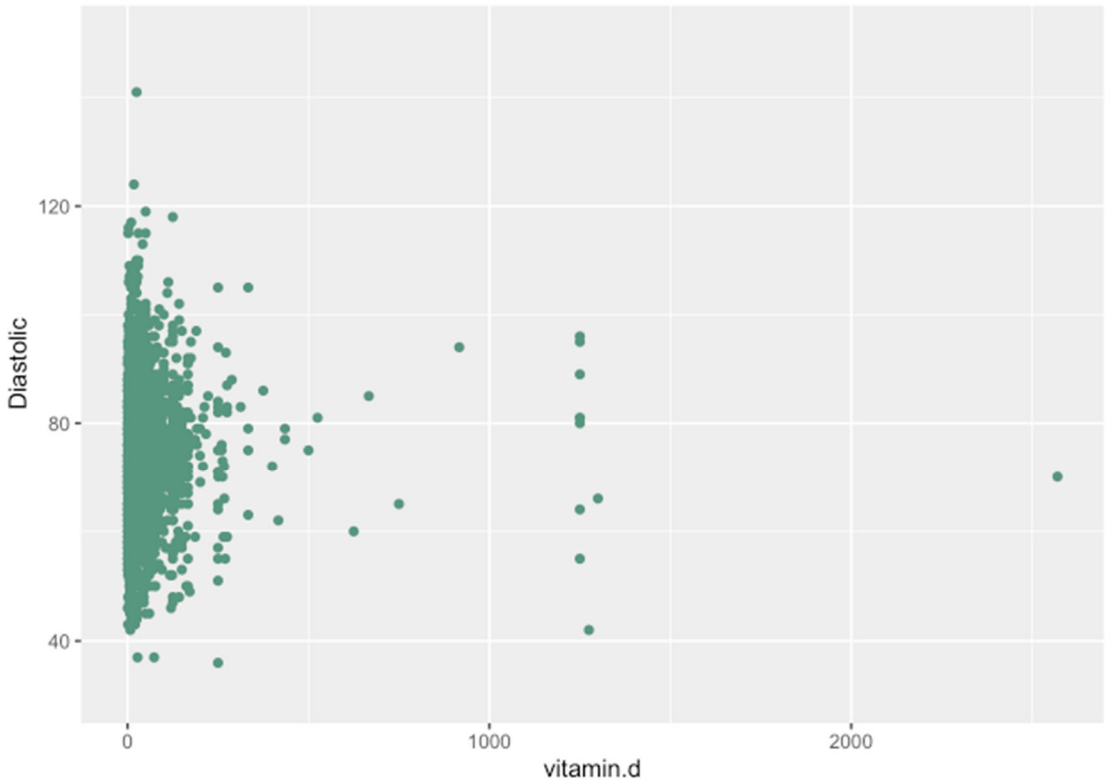
	elevated	high	normal	<NA>	Sum
Less than 9th grade	92	245	181	148	666
9-11th grade	152	343	338	114	947
High school graduate	366	702	790	202	2060
Some college	519	841	1173	238	2771
College graduate or above	423	548	930	186	2087
Refused	0	0	1	1	2
Don't Know	1	2	1	7	11
<NA>	212	52	2440	408	3112
Sum	1765	2733	5854	1304	11656

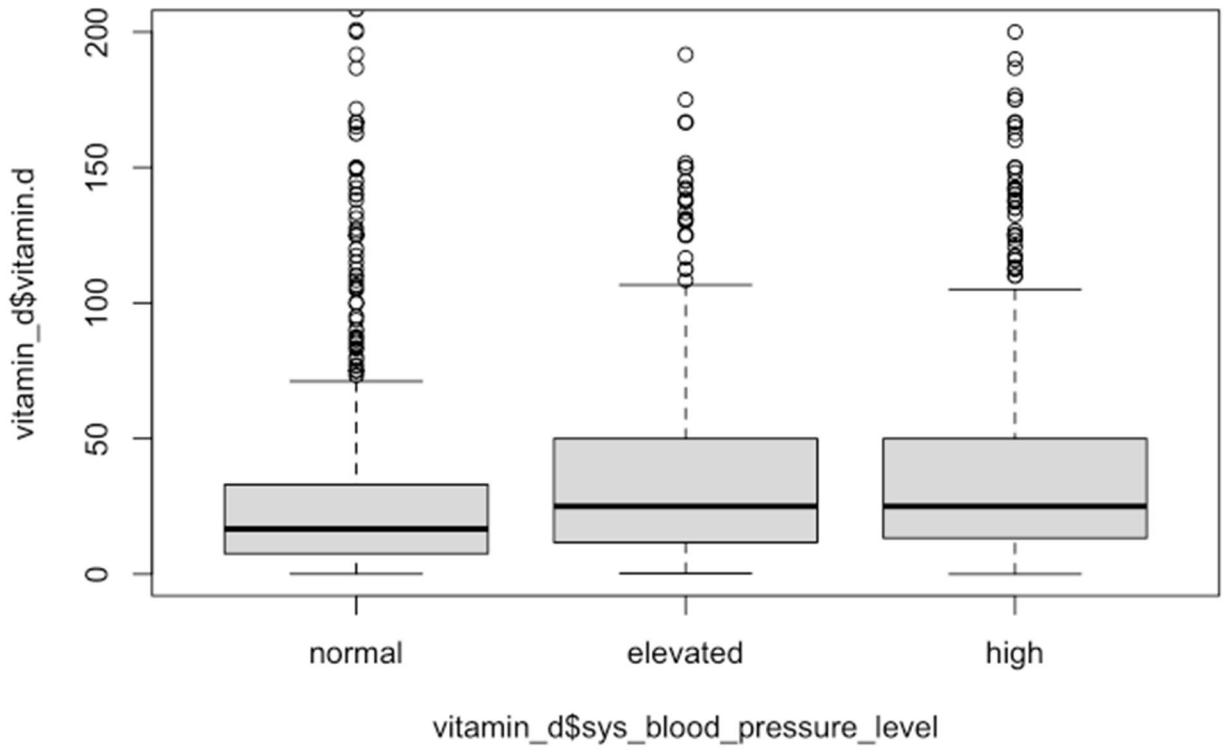
	high	normal	<NA>	Sum
Less than 9th grade	168	350	148	666
9-11th grade	279	554	114	947
High school graduate	641	1217	202	2060
Some college	892	1641	238	2771
College graduate or above	564	1337	186	2087
Refused	0	1	1	2
Don't Know	0	4	7	11
<NA>	105	2599	408	3112
Sum	2649	7703	1304	11656

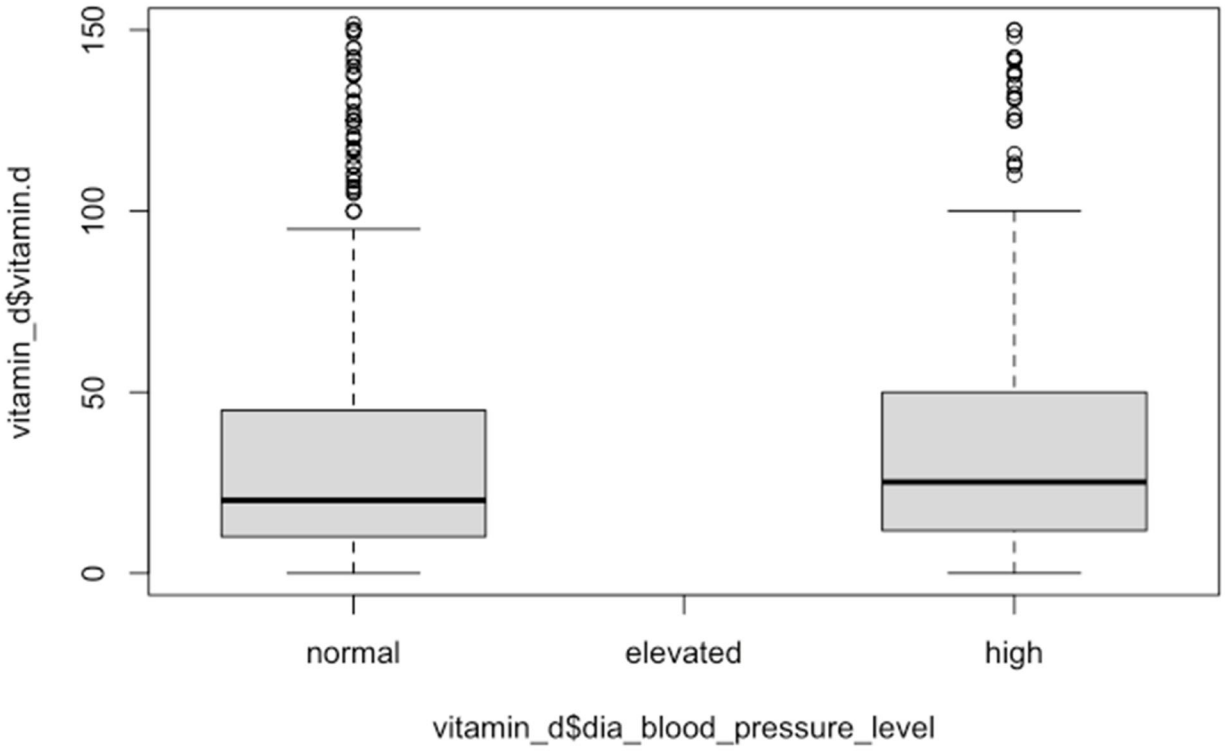
The next variable to explore is education. Do people with higher education have lower blood pressure? According to the bar charts and tables, participants who did not provide their education information have the best numbers – only 8.48% in elevated and high systolic blood pressures and 78.41% in normal systolic blood pressure, only 3.37% in high diastolic blood pressure and 83.52% in normal diastolic. If Refused, Don't know, and NA groups are excluded, then people with 9-11th grade education have the largest proportion in elevated and high systolic blood pressure. College graduate or above has the largest proportion in normal systolic blood pressure. For diastolic blood pressure, those with some college have the biggest proportion in the high blood pressure level. College graduate or above again has the largest proportion in the normal range.

Vitamin D vs. Blood Pressures






vitamin_d$vitamin.d
0
50
100
150
200
normal
elevated
high
vitamin_d$sys_blood_pressure_level


vitamin_d$vitamin.d
0
50
100
150
normal
elevated
high
vitamin_d$dia_blood_pressure_level

	sys_blood_pressure_level	Mean_vidamin_d	SD_vitamin_d
	<fct>	<dbl>	<dbl>
1	normal	32.7	60.1
2	elevated	51.0	140.
3	high	48.5	103.
4	<NA>	31.9	37.1

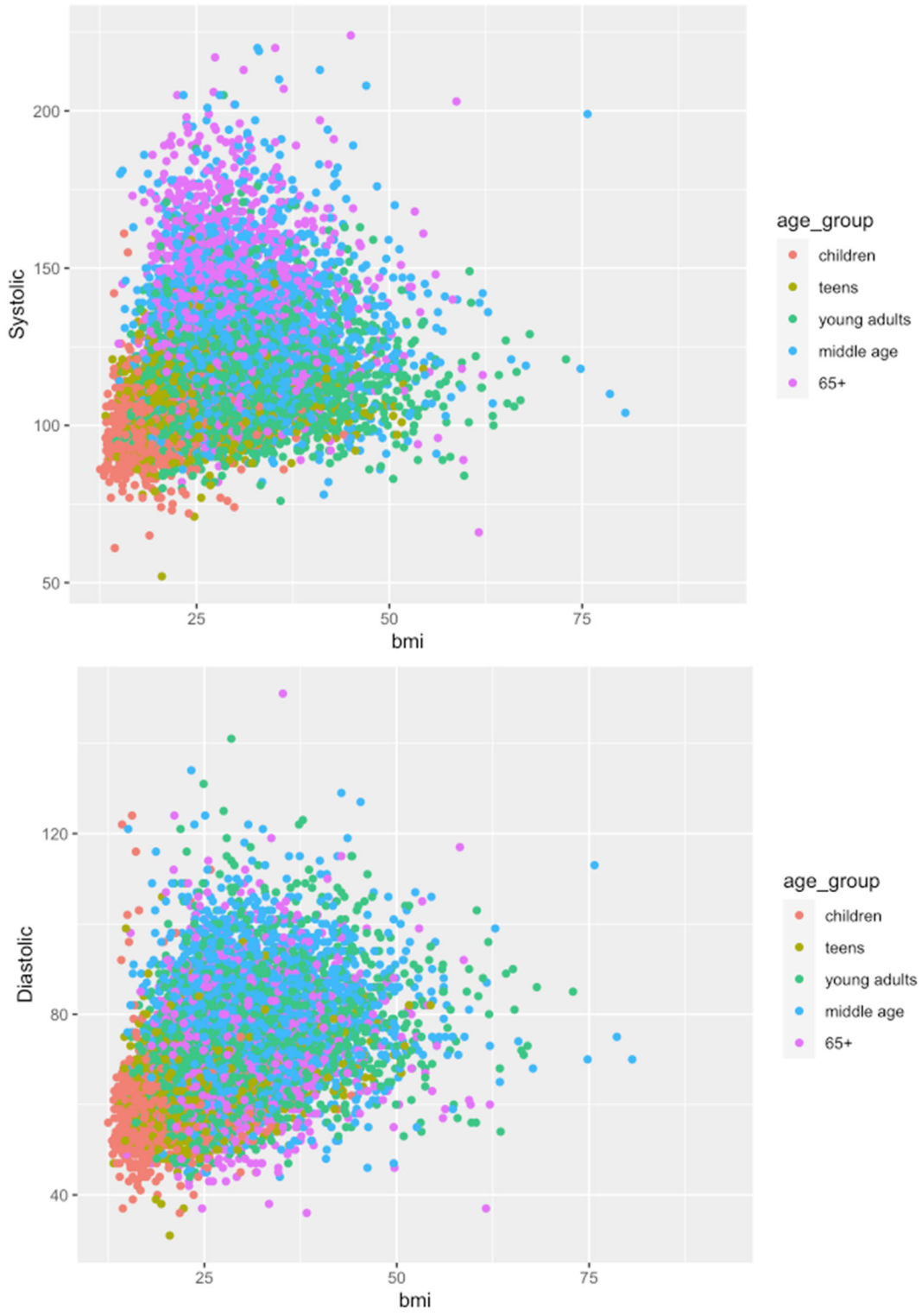
	dia_blood_pressure_level	Mean_vidamin_d	SD_vitamin_d
	<fct>	<dbl>	<dbl>
1	normal	37.9	85.7
2	high	50.5	114.
3	<NA>	31.9	37.1

The next variable to explore is daily Vitamin D dietary supplement intake. Is a high daily Vitamin D consumption associated with low blood pressure? The scatter plots do not demonstrate any obvious trends between Vitamin D supplement intake and blood pressures. However, box plots suggest that there may be a direct relationship. For both kinds of blood pressures, people with normal levels ingest less Vitamin D supplement. Normal systolic blood pressure group has a mean of 32.7 mcg and a standard deviation of 60.1 mcg Vitamin D intake. The elevated group has a mean of 51 mcg and a standard deviation of 140 mcg. The high blood pressure group has a mean of 48.5 mcg and a standard deviation of 103 mcg. Normal diastolic blood pressure group has a mean of 37.9 mcg and a standard deviation of 85.7 mcg Vitamin D intake, while the high blood pressure group has a mean of 50.0 mcg and a standard deviation of 114 mcg.

Body Mass Index vs. Blood Pressures

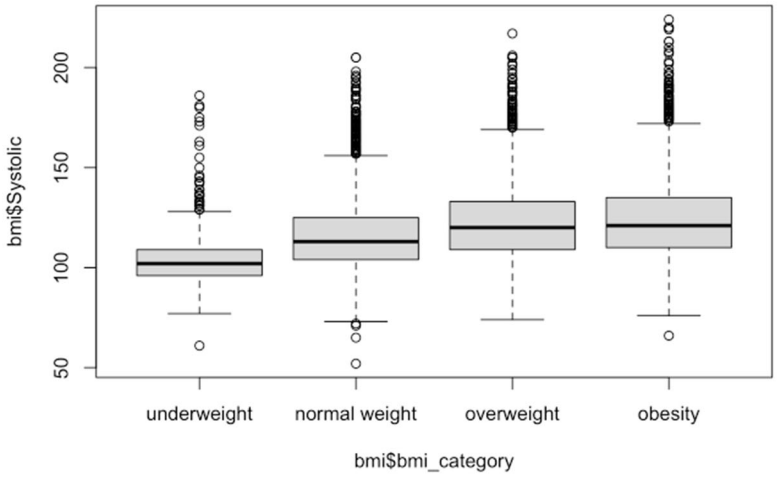

Systolic
200
150
100
50
25
50
75
bmi
gender_description
female
male

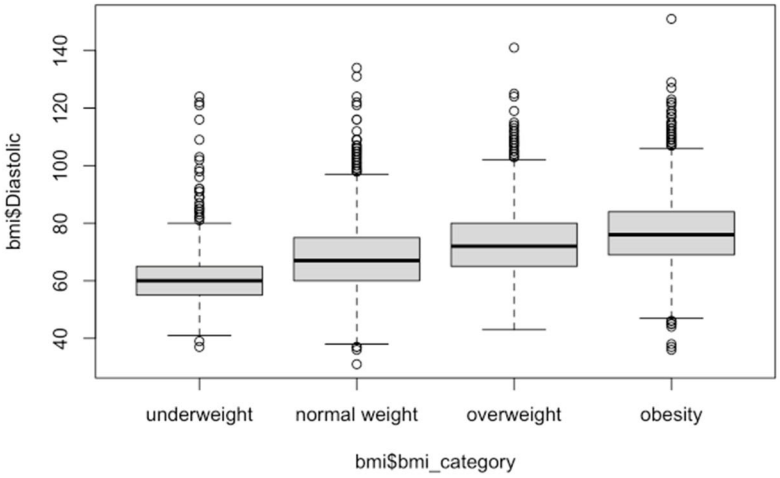

Diastolic
120
80
40
25
50
75
bmi
gender_description
female
male



Here, the relationships between Body Mass Index and blood pressures are examined. Both of the first two scatter plots show a clear direct trend - as BMI increases, blood pressures also increase. However, the fan-like shape

indicates heteroscedasticity, which may be troublesome for standard errors. The following two sets of scatter plots explore for any hidden trends that may surface with the incorporation of the gender or the age group variable. The direct relationship between BMI and blood pressure measurements appears to be true for all groups. But, the BMI variable displays heteroscedasticity in all distributions.





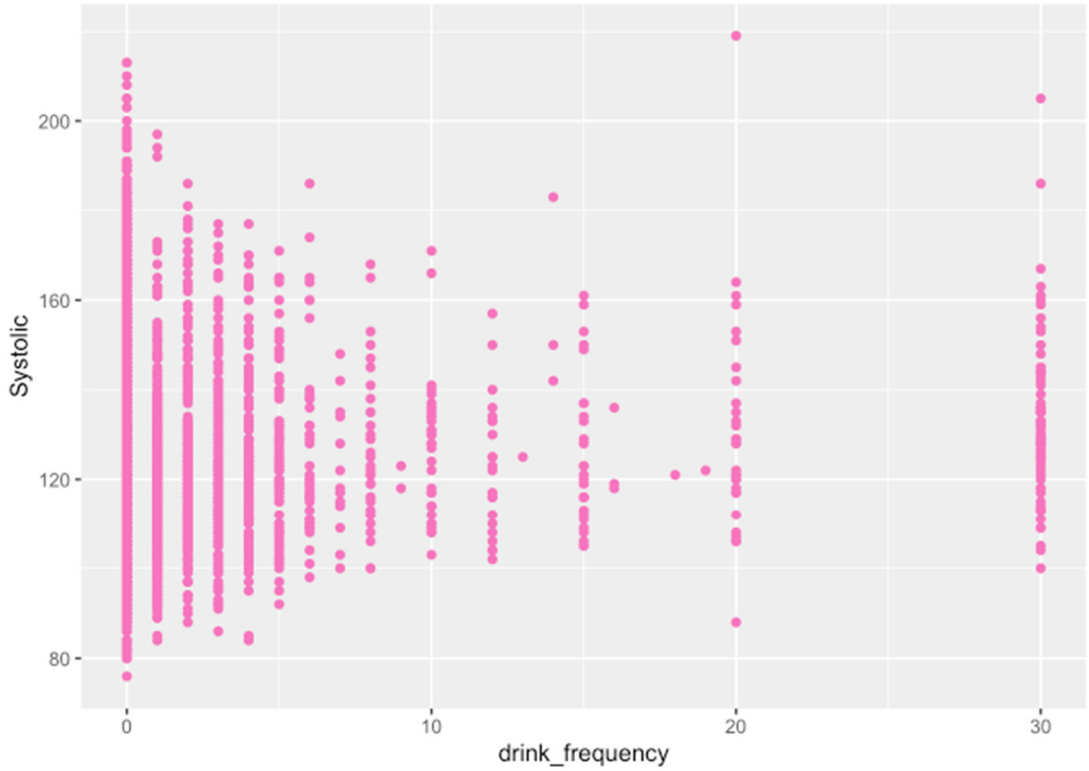
In order to work around the issue, a categorical variable called BMI category is created to separate BMI numbers into four groups: underweight, normal weight, overweight, and obesity. And, box plots are used to present the relationships. It is apparent that those with a lower BMI have lower blood pressures. The systolic means for the four BMI categories are: underweight is 103.6, normal weight is 116.6, overweight is 123.1, and obesity is 123.8 mmHg. The diastolic means are: underweight is 61.22, normal weight is 68.12, overweight is 73.07, and obesity is 76.87 mmHg. And, the higher the BMI, the more spread-out the blood pressure distributions are. The systolic interquartile ranges confirm such trait: underweight is 96-109, normal weight is 104-125,

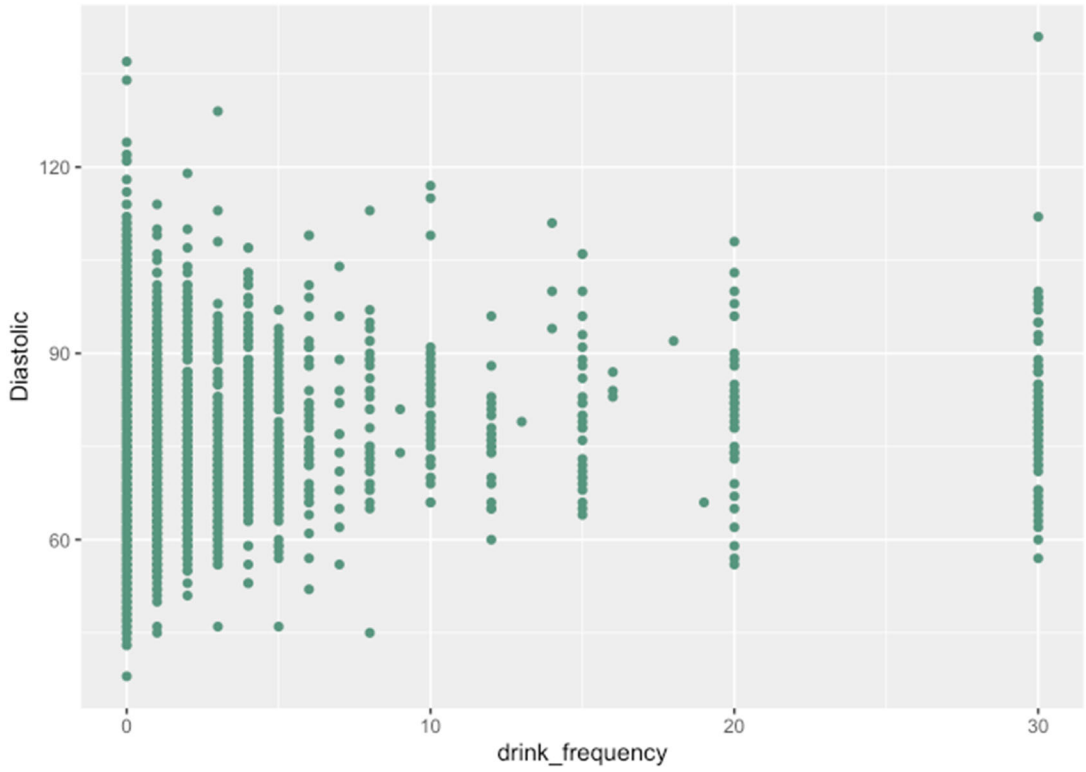
overweight is 109-133, and obesity is 110-135 mmHg. So are the diastolic interquartile ranges: underweight is 55-65, normal weight is 60-75, overweight is 65-80, and obesity is 69-84 mmHg.

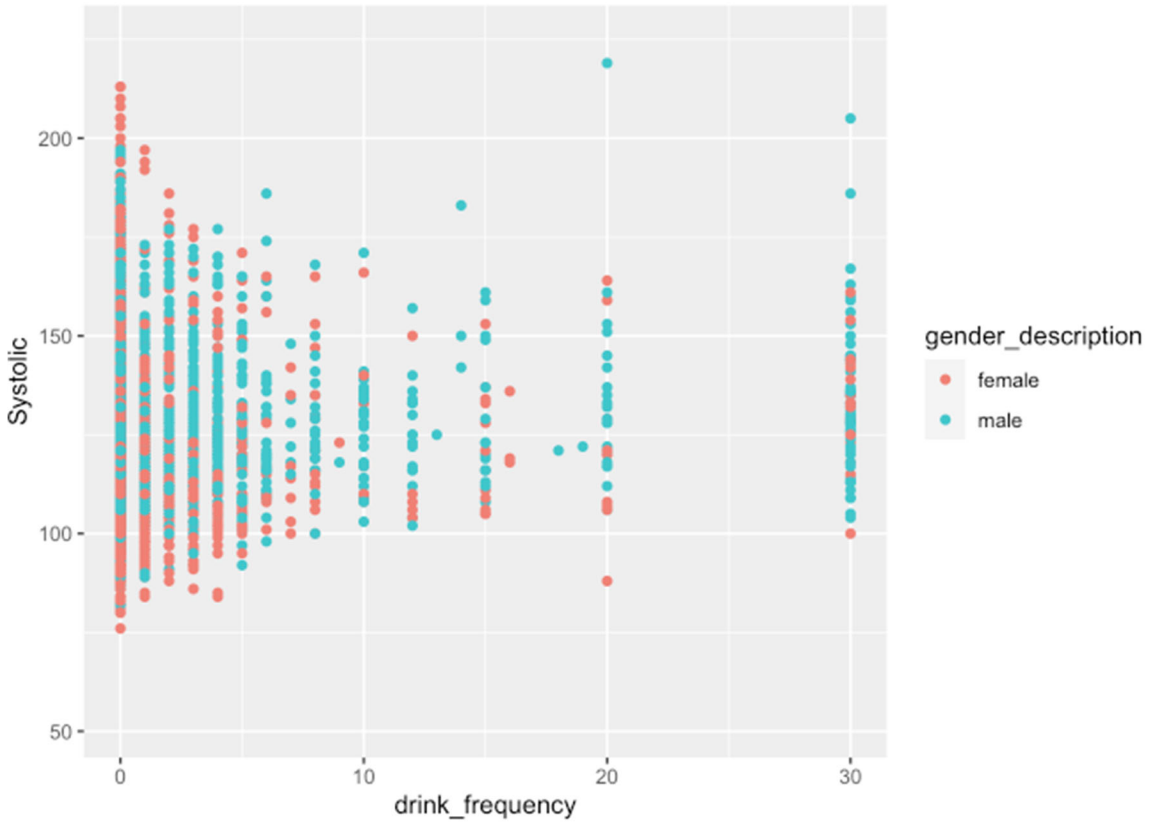
```
$systolic
$systolic$name
[1] "underweight"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   61.0    96.0   102.0   103.6   109.0   186.0     353

$systolic$name
[1] "normal weight"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   52.0   104.0   113.0   116.6   125.0   205.0     565

$systolic$name
[1] "overweight"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   74.0   109.0   120.0   123.1   133.0   217.0     537

$systolic$name
[1] "obesity"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   66.0   110.0   121.0   123.8   135.0   224.0     569
```

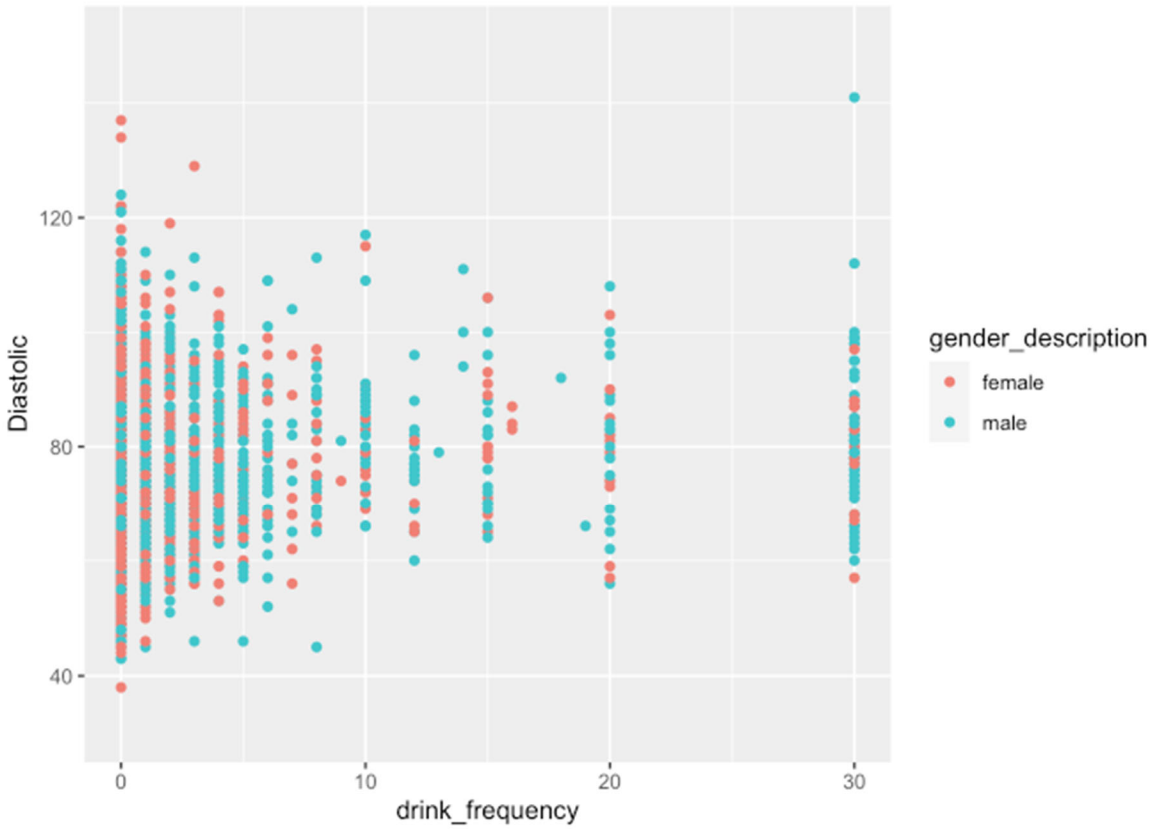
```
$diastolic
$diastolic$name
[1] "underweight"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  37.00   55.00   60.00   61.22   65.00  124.00     353

$diastolic$name
[1] "normal weight"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  31.00   60.00   67.00   68.12   75.00  134.00     565

$diastolic$name
[1] "overweight"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  43.00   65.00   72.00   73.07   80.00  141.00     537

$diastolic$name
[1] "obesity"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  36.00   69.00   76.00   76.87   84.00  151.00     569
```

Alcohol Consumption Frequency vs. Blood Pressures

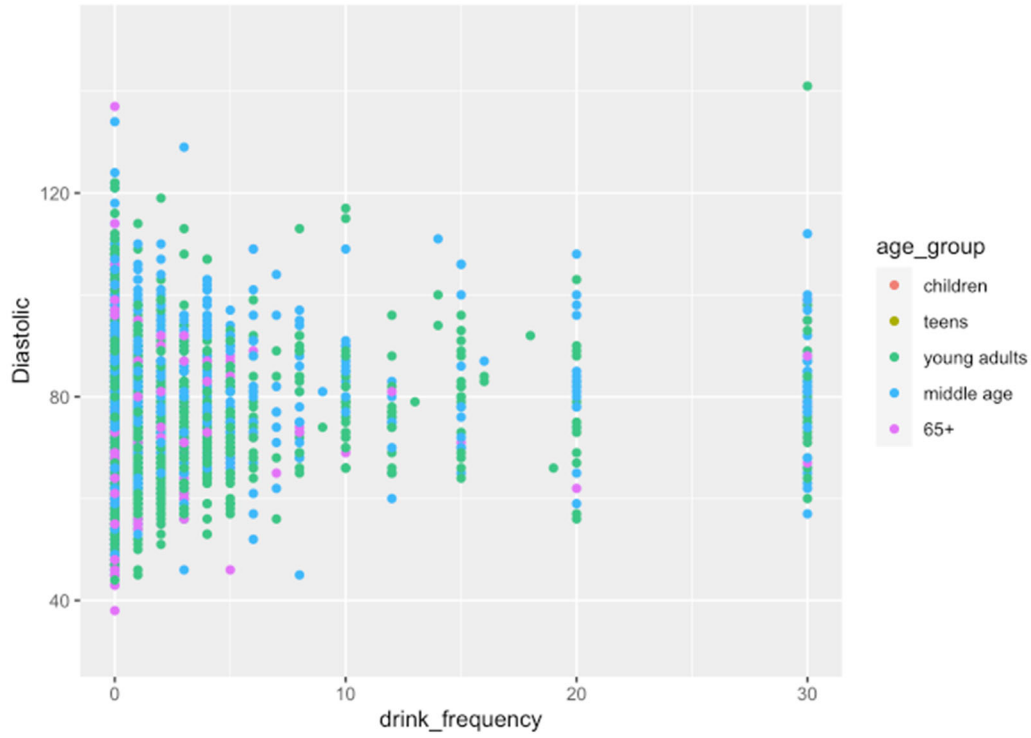


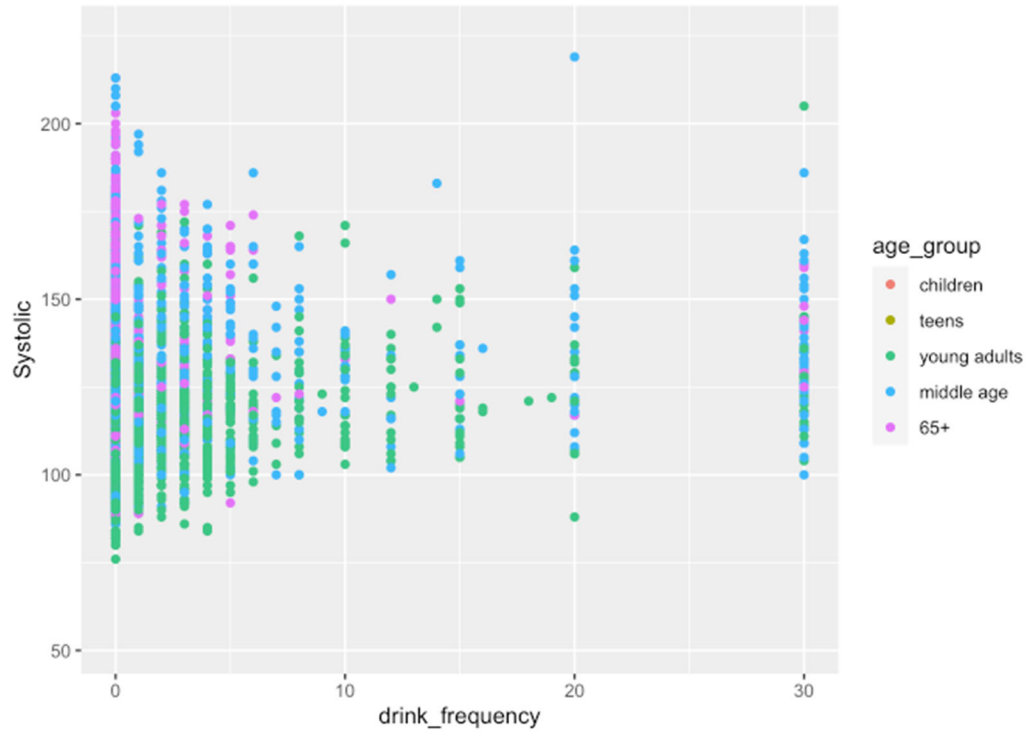



Systolic
200
150
100
50
0
10
20
30
drink_frequency
gender_description
female
male


Diastolic
120
80
40
0
10
20
30
drink_frequency
gender_description
female
male

	gender_description	Mean_alc	SD_alc	Mean_sys	SD_sys	Mean_dia	SD_dia
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	female	0.899	3.18	118.	20.8	71.6	12.2
2	male	1.73	4.88	122.	18.8	72.5	12.6

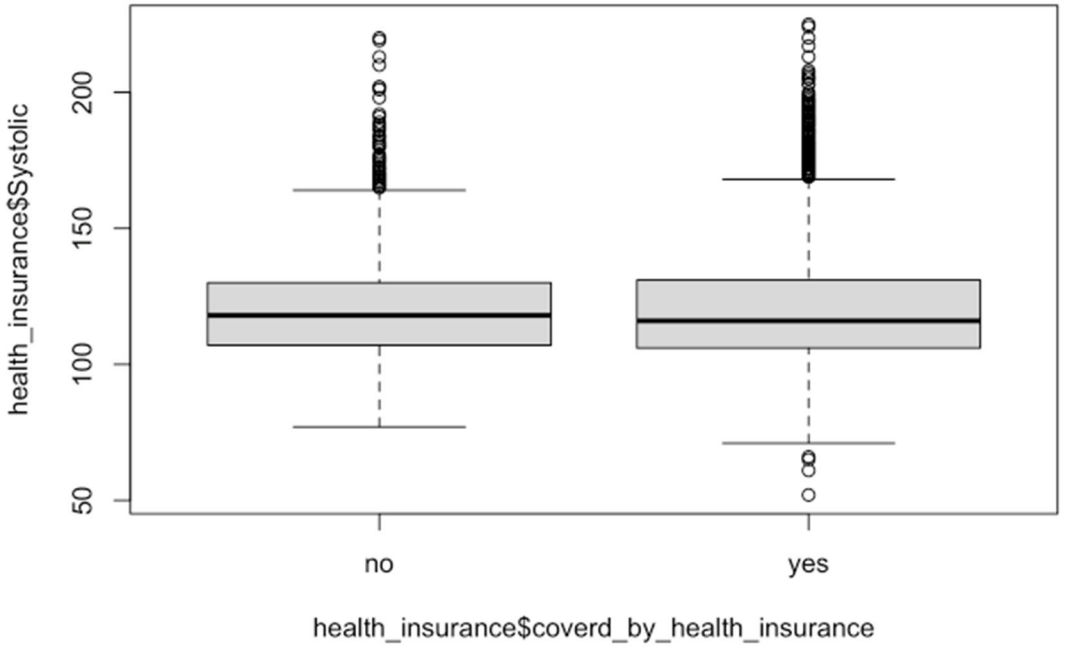


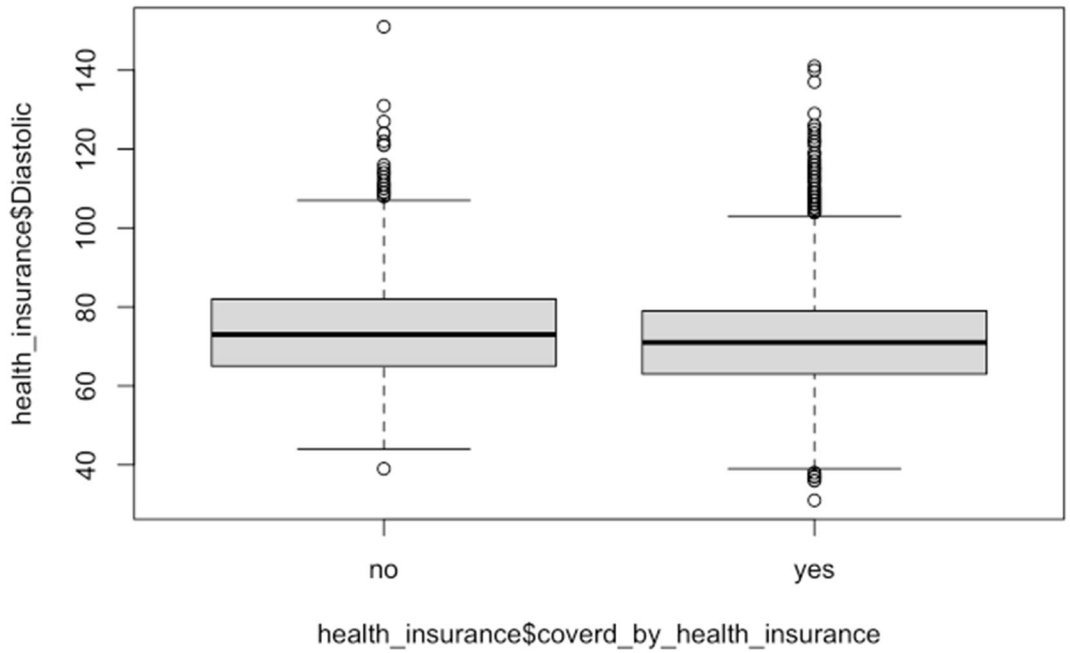


	age_group	Mean_alc	SD_alc	Mean_sys	SD_sys	Mean_dia	SD_dia
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	children	NaN	NA	101.	9.89	60.1	8.71
2	teens	NaN	NA	108.	10.4	64.4	8.30
3	young adults	1.44	3.87	115.	14.0	73.0	11.4
4	middle age	1.54	4.88	127.	19.1	78.3	11.6
5	65+	0.608	3.18	136.	21.8	72.9	12.1

Participants were asked the number of times they had 5 or more drinks on an occasion during the past 30 days. Is a high alcohol consumption frequency associated with high blood pressure? Scatter plots that examine the relationships between alcohol consumption and blood pressures do not indicate there are any trends. After incorporating the gender variable, there are still not obvious trends. Summary statistics show that on average, female drinks less than male (means: 0.899 vs. 1.73 times/30days, standard deviations: 3.18 vs. 4.88 times/30days) and have lower blood pressures (systolic means: 118 vs. 122 mmHg, standard deviations: 20.8 vs. 18.8 mmHg. Diastolic means: 71.6 vs. 72.5 mmHg, standard deviations: 12.2 vs. 12.6 mmHg). When the interaction of drink frequency and age group is explored, only the age group of 65 and above shows an inverse relationship between alcohol consumption and systolic blood pressure. The rest of the trends seem to be flat. Compared to young adults and middle-age adults, people who are 65 and above drink much less on average (0.61 vs. 1.44 vs. 1.54 times/30days), have the lowest average diastolic blood pressure (72.9 vs. 73 vs. 78.3 mmHg) but have the highest systolic blood pressure (136 vs. 115 vs. 127 mmHg).

Health Insurance vs. Blood Pressure



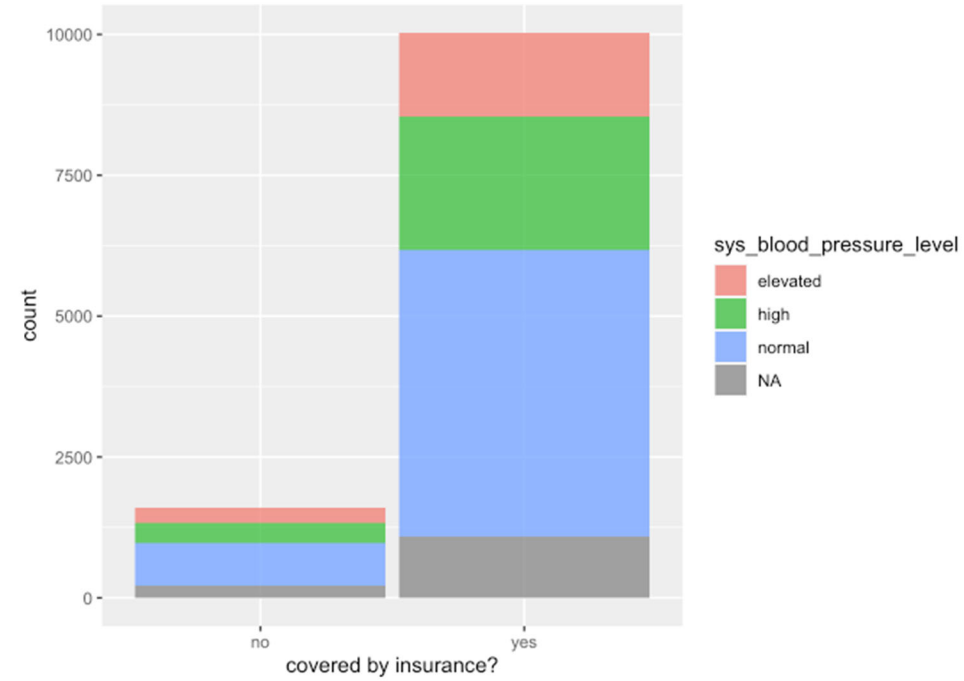


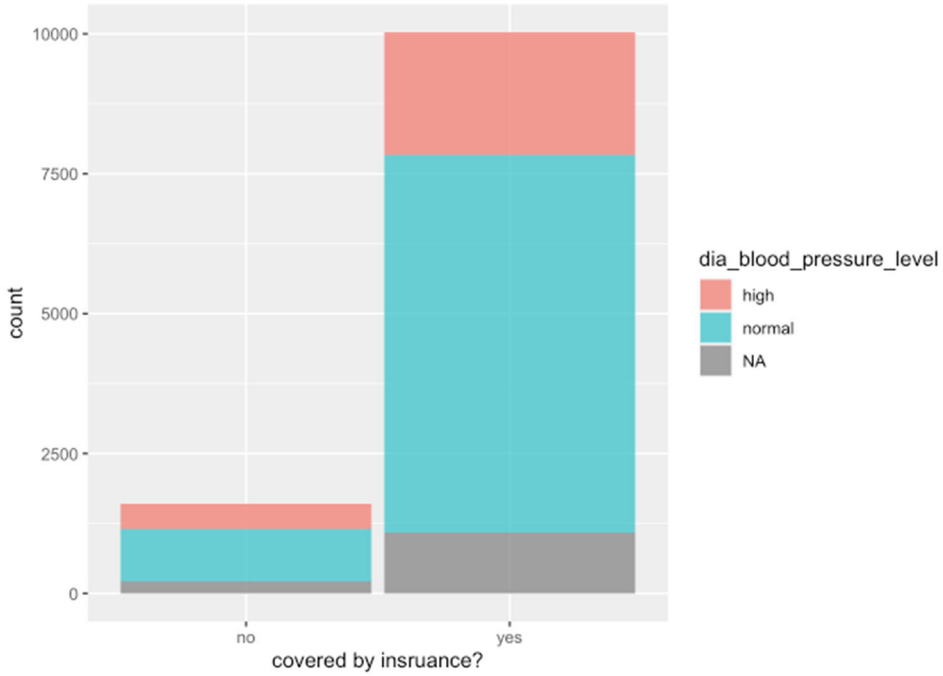
```
$systolic
$systolic$name
[1] "yes"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   52.0   106.0   116.0   119.8   131.0   225.0    1083

$systolic$name
[1] "no"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   77.0   107.0   118.0   121.1   130.0   220.0     219
```

```
$diastolic
$diastolic$name
[1] "yes"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  31.00   63.00   71.00   71.65   79.00  141.00    1083

$diastolic$name
[1] "no"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   39.0    65.0    73.0    74.5    82.0   151.0     219
```

Do people who have health insurance generally have lower blood pressure? Based on box plots and summary statistics, the difference seems to be negligible. The systolic blood pressure mean for those who have insurance coverage is 119.8 mmHg, and the center 50% data points fall in the 106-131 mmHg range. The mean for those who do not have insurance is 121.1 mmHg, and the center 50% data points are in the 107-130 mmHg range. For diastolic blood pressure, those who answered “yes” to insurance coverage have a mean of 71.65 mmHg and an interquartile range of 63-79 mmHg. Those who answered “no” have a mean of 74.5 mmHg and an interquartile range of 64-82 mmHg.

The bar charts and tables show that overall, people without insurance coverage have higher proportions in the elevated and high blood pressure levels for both of systolic and diastolic – no’s 39.49% vs. yes’ 38.47% for systolic, and no’s 28.1% vs. yes’ 21.87% for diastolic. They also have lower proportions in the normal blood pressure levels – no’s 46.81% vs. yes’ 50.73% for systolic, and no’s 58.2% vs. yes’ 67.33% for diastolic.

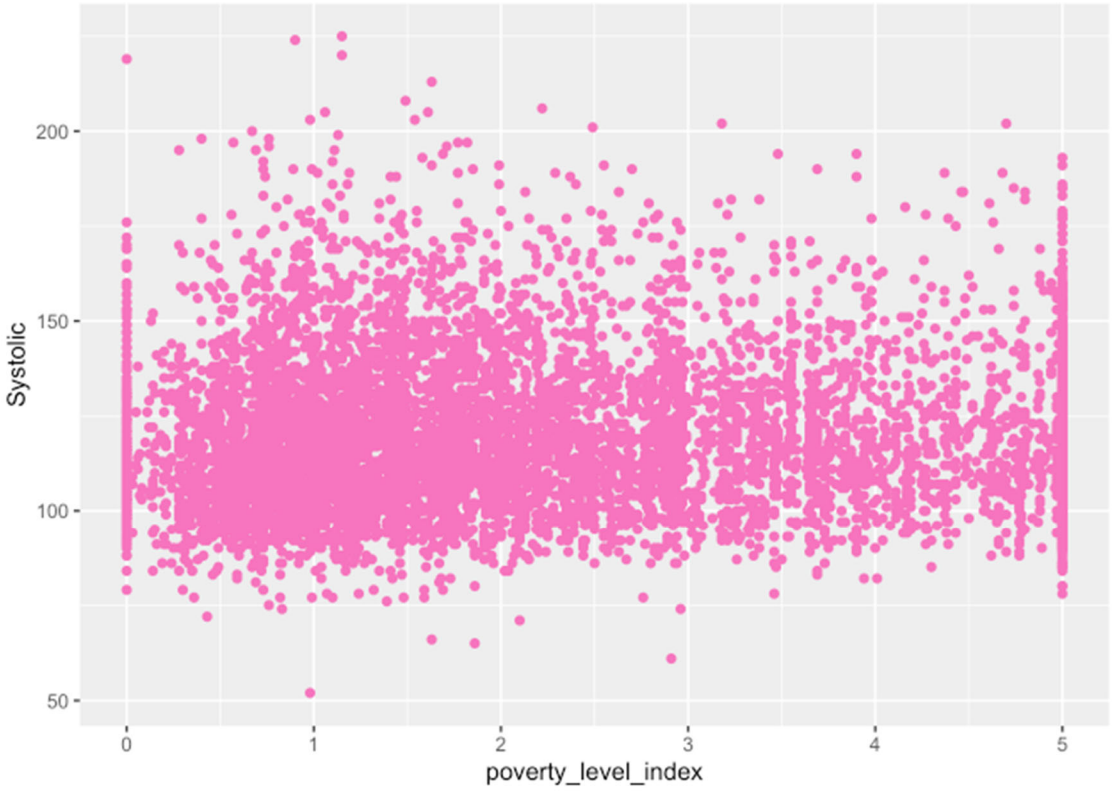


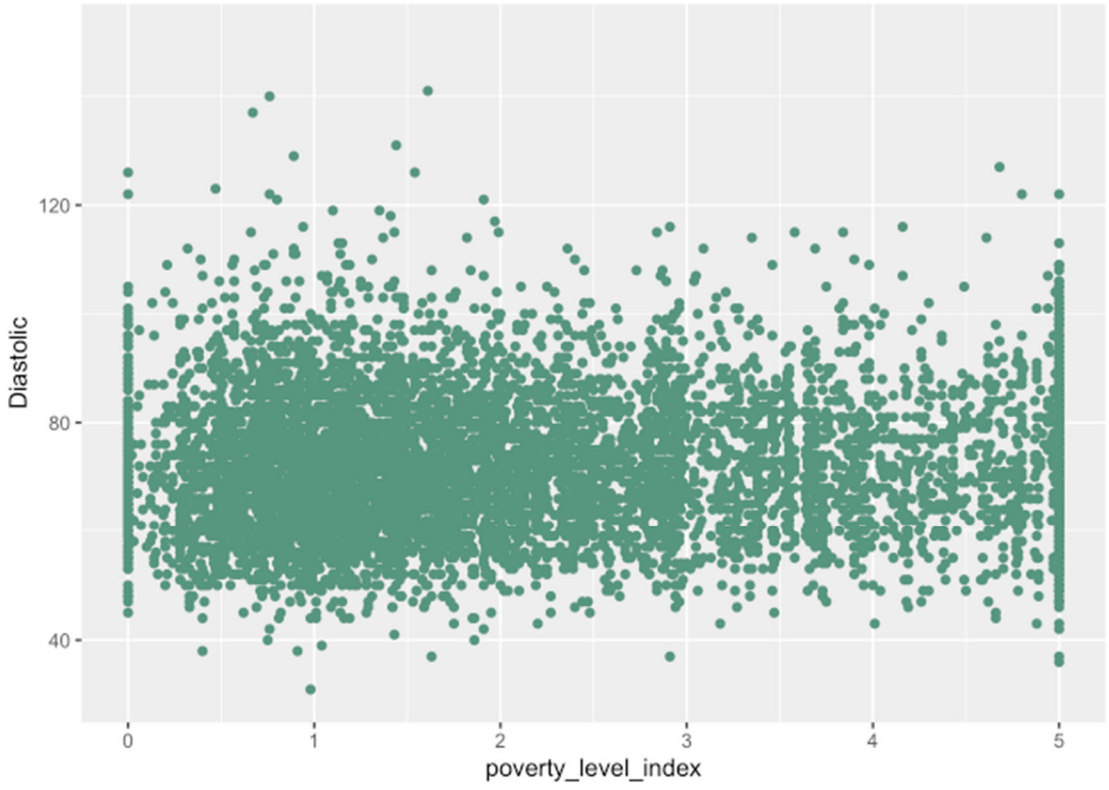


	elevated	high	normal	<NA>	Sum
no	273	358	748	219	1598
yes	1487	2372	5089	1083	10031
<NA>	0	0	0	0	0
Sum	1760	2730	5837	1302	11629

	high	normal	<NA>	Sum
no	449	930	219	1598
yes	2194	6754	1083	10031
<NA>	0	0	0	0
Sum	2643	7684	1302	11629

Income vs. Blood Pressures





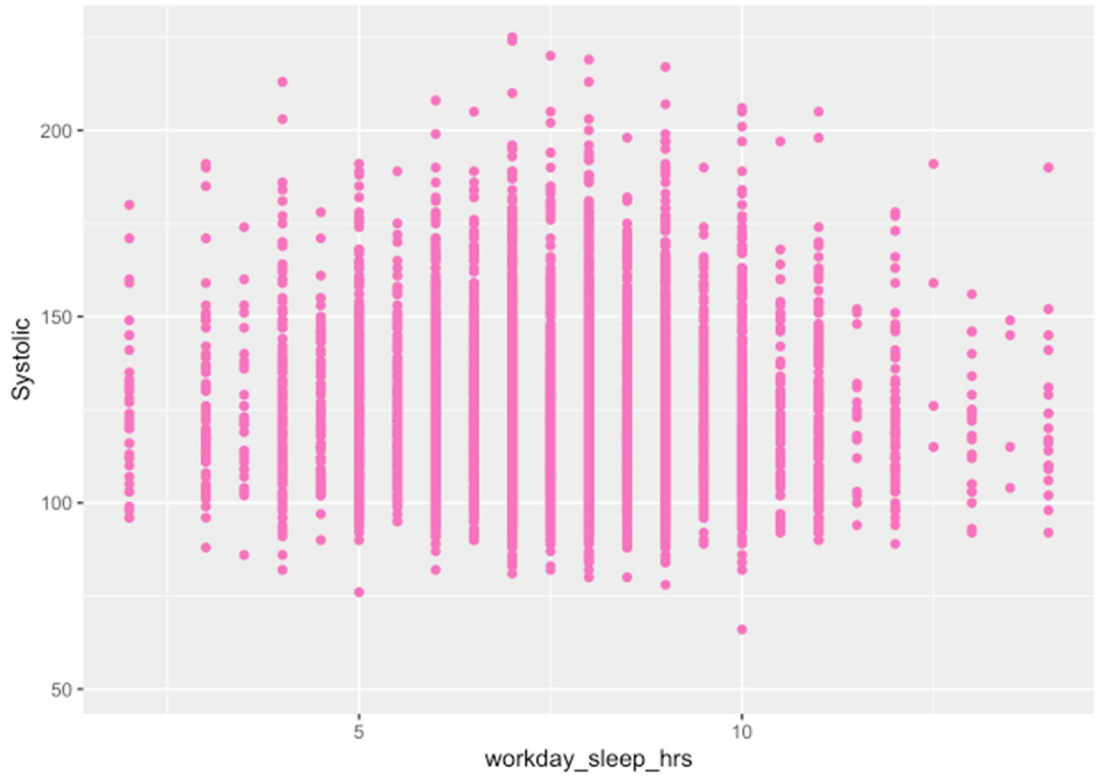
Mean_in <dbl>	SD_inc <dbl>	Min_inc <dbl>	Max_inc <dbl>
2.259799	1.549095	0	5

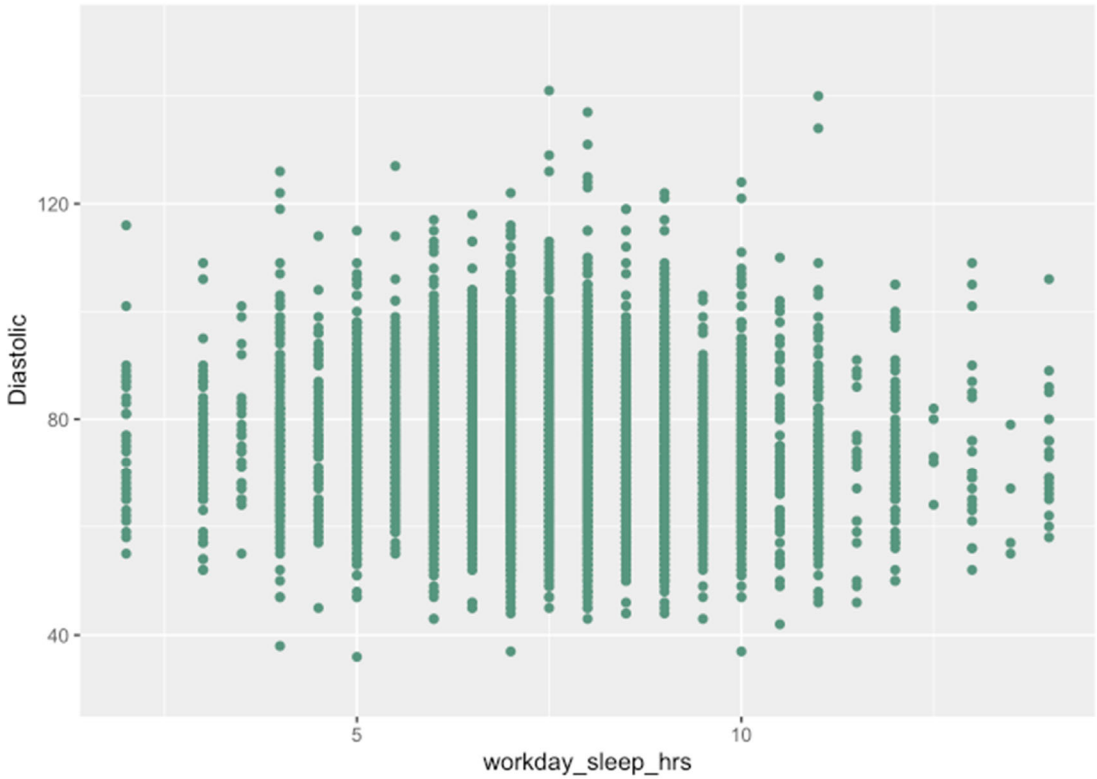
Mean_sys <dbl>	SD_sys <dbl>	Min_sys <int>	Max_sys <int>
120.0046	19.95013	52	225

Mean_dia <dbl>	SD_dia <dbl>	Min_dia <int>	Max_dia <int>
72.03574	12.41688	31	151

Next, the variable, income, is examined. The respondents were asked to report total family income, received last month in dollars, for themselves and the other members of their family. This variable is an index for the ratio of reported monthly income to poverty. The Department of Health and Human Services' (HHS) poverty guidelines were used as the poverty measure to calculate the index. My question was: is high income associated with low blood pressure? The scatter plots suggest, although not obvious, there may be an inverse relationship for both income vs. systolic blood pressure and income vs. diastolic blood pressure. The mean for the poverty level index variable is 2.26 with a standard deviation of 1.55. Minimum is 0, and maximum is 5.

Sleep vs. Blood Pressures





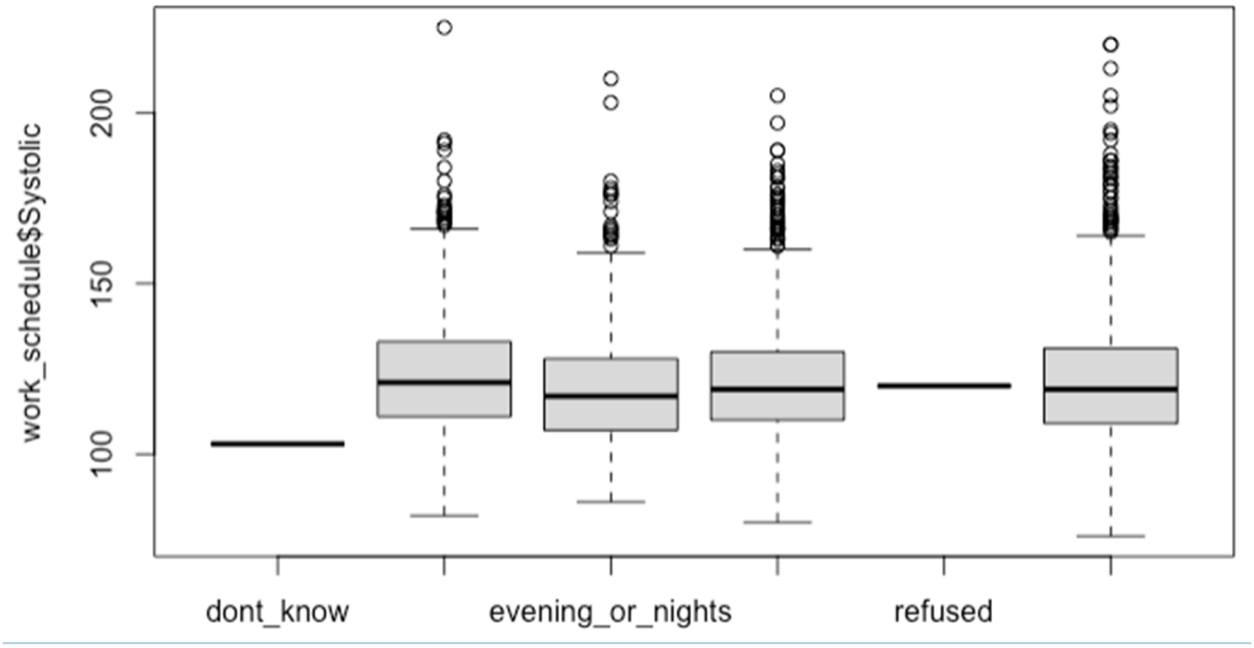
Mean_sleep <dbl>	SD_sleep <dbl>	Min_sleep <dbl>	Max_sleep <dbl>
7.625908	1.682205	2	14

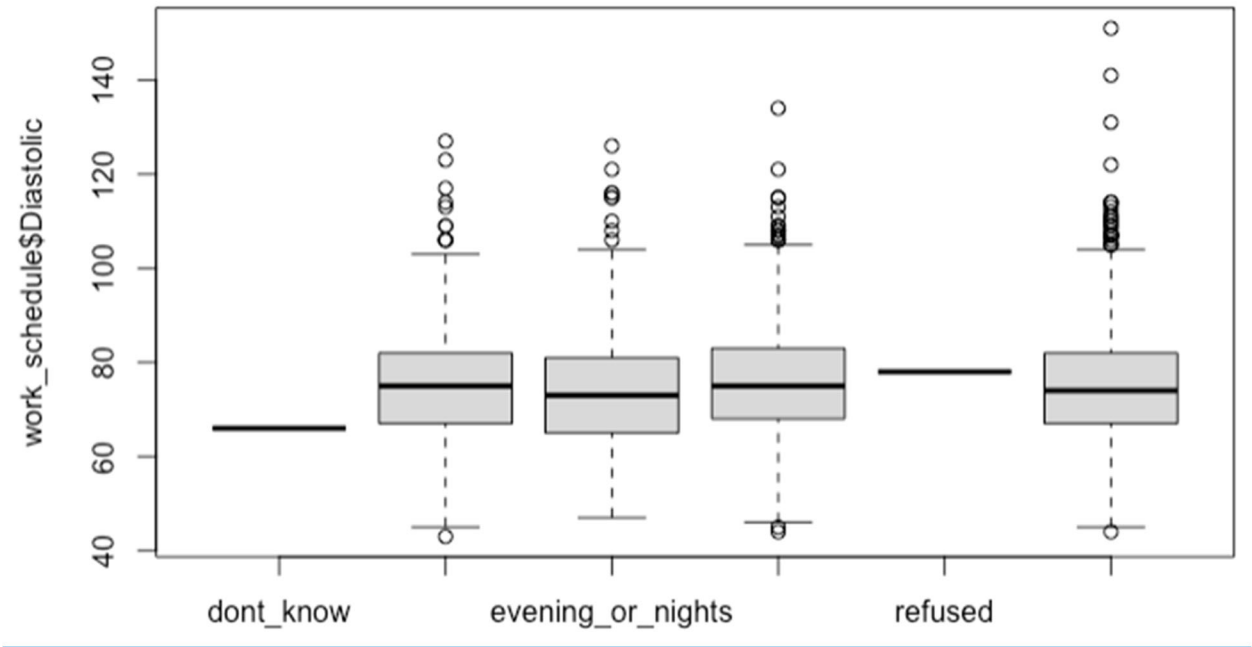
Mean_sys <dbl>	SD_sys <dbl>	Min_sys <int>	Max_sys <int>
120.0046	19.95013	52	225

Mean_dia <dbl>	SD_dia <dbl>	Min_dia <int>	Max_dia <int>
72.03574	12.41688	31	151

Human subjects were asked the number of hours they usually sleep on weekdays or workdays. Therefore, the next question to explore is: is little sleep associated with high blood pressure? The scatter plots suggest, although not obvious, there may be an inverse relationship for both number of hours of sleep vs. systolic blood pressure and number of hours of sleep vs. diastolic blood pressure. The mean for the sleep variable is 7.63 hours, and the standard deviation is 1.68 hours. Minimum is 2, and maximum is 14.

Work Schedule vs. Blood Pressures

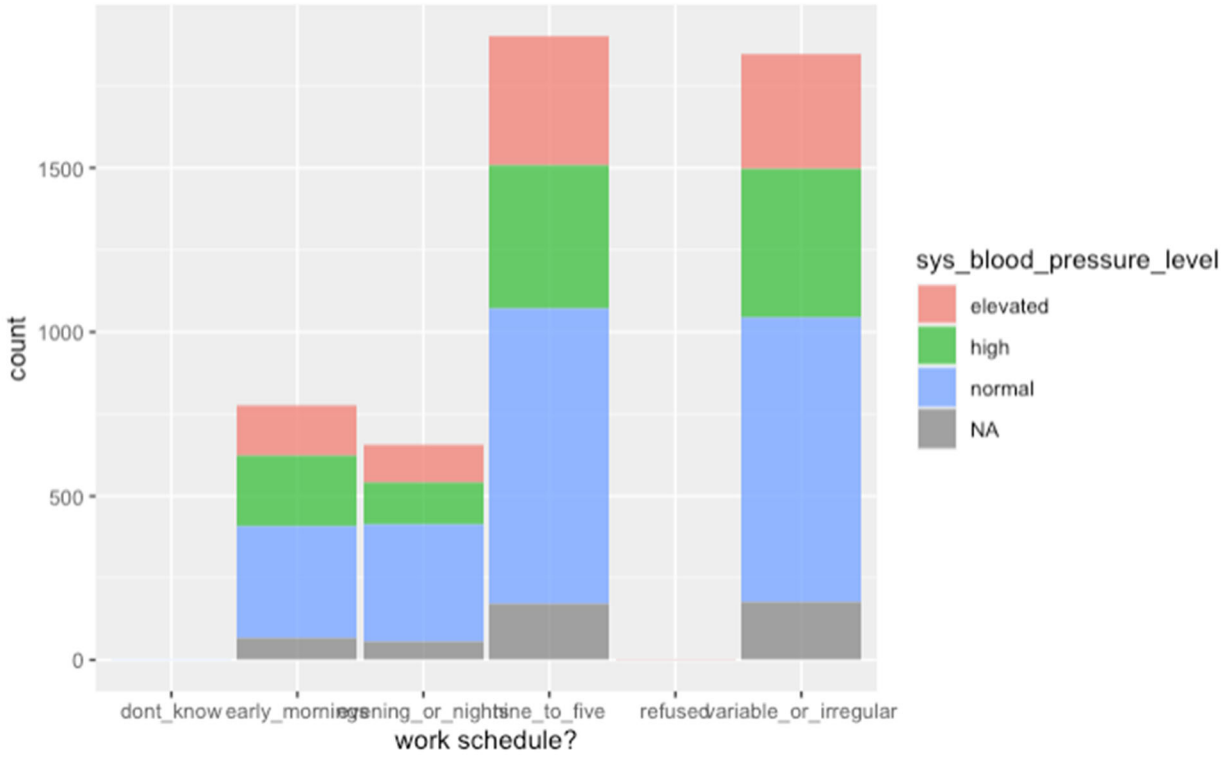


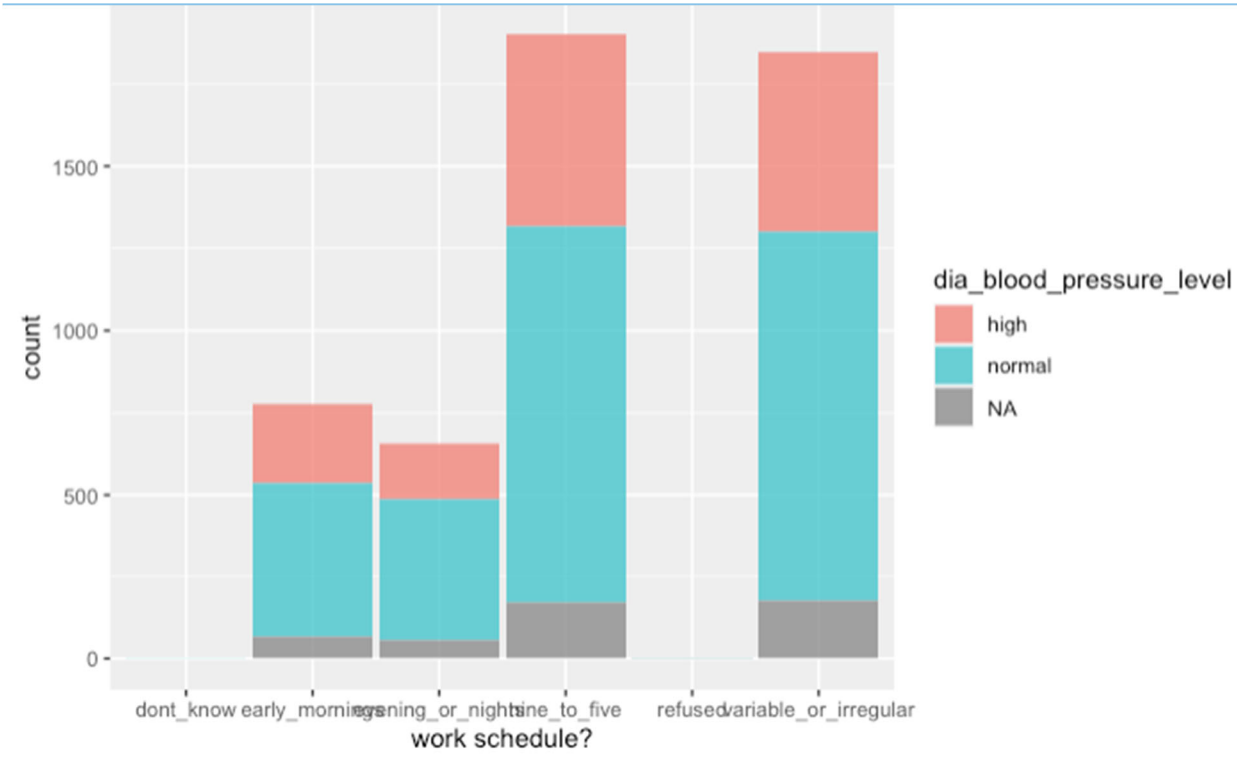


```
$systolic
$systolic$name
[1] "nine_to_five"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
     80     110     119     121     130     205     168

$systolic$name
[1] "evening_or_nights"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   86.0   107.0   117.0   119.1   128.0   210.0      55

$systolic$name
[1] "early_mornings"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   82.0   111.0   121.0   123.6   133.0   225.0      69

$systolic$name
[1] "variable_or_irregular"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   76.0   109.0   119.0   121.2   131.0   220.0     175

$systolic$name
[1] "refused"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    120     120     120     120     120     120       1

$systolic$name
[1] "dont_know"

$systolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    103     103     103     103     103     103
```

```
$diastolic
$diastolic$name
[1] "nine_to_five"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   44.0    68.0    75.0    75.6    83.0   134.0     168

$diastolic$name
[1] "evening_or_nights"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  47.00   65.00   73.00   73.63   81.00  126.00      55

$diastolic$name
[1] "early_mornings"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   43.0    67.0    75.0    75.5    82.0   127.0      69

$diastolic$name
[1] "variable_or_irregular"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  44.00   67.00   74.00   74.93   82.00  151.00     175

$diastolic$name
[1] "refused"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
     78      78      78      78      78      78       1

$diastolic$name
[1] "dont_know"

$diastolic$summary_statistic
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     66      66      66      66      66      66
```

Lastly, how is one's work schedule related to blood pressures? Box plots suggest participants who have an early-morning schedule may have a slightly higher systolic blood pressure than the rest of the groups. And those who have an evening/nights work schedule may have the lowest systolic and diastolic blood pressures. As for diastolic blood pressure, all the work schedules share a similar distribution.


count
1500
1000
500
0
dont_know
early_morning
evening_or_night
nine_to_five
refused
variable_or_irregular
work schedule?
sys_blood_pressure_level
elevated
high
normal
NA


count
1500
1000
500
0
dont_know
early_morning
evening_or_night
nine_to_five
refused
variable_or_irregular
work schedule?
dia_blood_pressure_level
high
normal
NA

	elevated	high	normal	<NA>	Sum
dont_know	0	0	1	0	1
early_mornings	151	218	336	69	774
evening_or_nights	115	133	356	55	659
nine_to_five	393	439	903	168	1903
refused	1	0	0	1	2
variable_or_irregular	349	450	871	175	1845
<NA>	0	0	0	0	0
Sum	1009	1240	2467	468	5184

	high	normal	<NA>	Sum
dont_know	0	1	0	1
early_mornings	238	467	69	774
evening_or_nights	174	430	55	659
nine_to_five	586	1149	168	1903
refused	0	1	1	2
variable_or_irregular	546	1124	175	1845
<NA>	0	0	0	0
Sum	1544	3172	468	5184

To explore the relationships in another way, bar charts and count tables are generated. If those who reported “don’t know” or “refused” are excluded, then early-morning workers have the biggest proportion in the elevated and high systolic blood pressure category – 47.67%, and evening/nights workers have the highest proportion in the normal systolic blood pressure category – 54.02%. Nine-to-five workers have the largest proportion in high diastolic blood pressure – 30.79%, and evening/nights workers again have the highest proportion in the normal diastolic blood pressure category – 65.25%.

Summary of Findings, Potential Statistical Analysis for Part II

In summary, both of one's systolic and diastolic blood pressures do not fluctuate too much under general conditions. The two kinds of blood pressures are highly correlated and share a direct relationship. Female may have significantly lower blood pressures than male. One's age may also be directly related to their systolic blood pressure, as well as their diastolic blood pressure if the person is before 50 years old. After 50, the person is likely to have lower diastolic blood pressure compared to younger adults. If a person is African American, they will likely experience higher blood pressures than the other races. If a person is Mexican American, they may experience relatively lower blood pressures. People with 9-11th grade education level are most likely to be found to have elevated or high systolic blood pressure. Those with some college education are most likely to be found to have high diastolic blood pressure. Participants with college graduate or above education level have the largest proportions in normal blood pressures. Vitamin D supplement consumption wise, human subjects who have less Vitamin D supplements seem to have lower blood pressures. One's BMI is also directly related to blood pressures. To my surprise, for people who are 65 years old and above, alcohol consumption is inversely related to diastolic blood pressure, while there are not obvious trends between other age groups and blood pressures. Health insurance wise, those with health insurance coverage show higher proportions in the normal blood pressure range and smaller proportions in the elevated or hypertension range. There may be an inverse relationship between the poverty level index quantitative variable and blood pressures. The same goes for hours of sleep on workdays and blood pressures. Lastly, early-morning workers are most likely to have elevated or high systolic blood pressure, and nine-to-five workers are most likely to have high diastolic blood pressure. Evening/nights workers have the highest proportions in normal blood pressure ranges.

In Part II of the projects, I would like to carry out various tests to examine the observations above. Simple linear regression will be used to verify the direct relationship between the two types of blood pressure. The same test will be good for the relationships for blood pressures versus the following quantitative variables: age, BMI, poverty level index, and hours of sleep on workdays. ANOVA will be used to determine if the differences between female and male's blood pressures are significant. The same test will be used to examine the relationship between Vitamin D supplement intake and blood pressures. Multiple linear regression will be used to test the relationship between the interactions of alcohol consumption and age and blood pressures. Chi-square test will be good for comparing blood pressure level proportions divided by categories in the following variables: race, education level, health insurance coverage, and work schedule.