Process Followed:

The dataset chosen is that of diabetes and factors that may be affecting it. It contains medical records of patients with features like glucose level, BMI, blood pressure, etc., along with an Outcome column indicating whether a patient has diabetes.
It is used to analyze which health factors strongly or weakly influence diabetes prediction.

```
> str(df)
'data.frame':   768 obs. of  9 variables:
 $ Pregnancies              : int  6 1 8 1 0 5 3 10 2 8 ...
 $ Glucose                  : int  148 85 183 89 137 116 78 115 197
 $ BloodPressure            : int  72 66 64 66 40 74 50 0 70 96 ...
 $ SkinThickness            : int  35 29 0 23 35 0 32 0 45 0 ...
 $ Insulin                  : int  0 0 0 94 168 0 88 0 543 0 ...
 $ BMI                      : num  33.6 26.6 23.3 28.1 43.1 25.6 31
 $ DiabetesPedigreeFunction: num  0.627 0.351 0.672 0.167 2.288 ..
 $ Age                      : int  50 31 32 21 33 30 26 29 53 54 ..
 $ Outcome                  : int  1 0 1 0 1 0 1 0 1 1 ...
```

```
> summary(df)
  Pregnancies        Glucose      BloodPressure    SkinThickness
 Min.   : 0.000   Min.   :  0.0   Min.   :  0.00   Min.   : 0.00
 1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.00   1st Qu.: 0.00
 Median : 3.000   Median :117.0   Median : 72.00   Median :23.00
 Mean   : 3.845   Mean   :120.9   Mean   : 69.11   Mean   :20.54
 3rd Qu.: 6.000   3rd Qu.:140.2   3rd Qu.: 80.00   3rd Qu.:32.00
 Max.   :17.000   Max.   :199.0   Max.   :122.00   Max.   :99.00
    Insulin           BMI        DiabetesPedigreeFunction      Age
 Min.   :  0.0   Min.   : 0.00   Min.   :0.0780           Min.   :21.00
 1st Qu.:  0.0   1st Qu.:27.30   1st Qu.:0.2437           1st Qu.:24.00
 Median : 30.5   Median :32.00   Median :0.3725           Median :29.00
 Mean   : 79.8   Mean   :31.99   Mean   :0.4719           Mean   :33.24
 3rd Qu.:127.2   3rd Qu.:36.60   3rd Qu.:0.6262           3rd Qu.:41.00
 Max.   :846.0   Max.   :67.10   Max.   :2.4200           Max.   :81.00
    Outcome
 Min.   :0.000
 1st Qu.:0.000
 Median :0.000
 Mean   :0.349
 3rd Qu.:1.000
 Max.   :1.000
```

Calculated the correlation matrix

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Pregnancies	1.00000000	0.12945867	0.14128198	-0.08167177	-0.07353461	0.01768309	-0.03352267	0.54434123	0.22189815
Glucose	0.12945867	1.00000000	0.15258959	0.05732789	0.33135711	0.22107107	0.13733730	0.26351432	0.46658140
BloodPressure	0.14128198	0.15258959	1.00000000	0.20737054	0.08893338	0.28180529	0.04126495	0.23952795	0.06506836
SkinThickness	-0.08167177	0.05732789	0.20737054	1.00000000	0.43678257	0.39257320	0.18392757	-0.11397026	0.07475223
Insulin	-0.07353461	0.33135711	0.08893338	0.43678257	1.00000000	0.19785906	0.18507093	-0.04216295	0.13054795
BMI	0.01768309	0.22107107	0.28180529	0.39257320	0.19785906	1.00000000	0.14064695	0.03624187	0.29269466
DiabetesPedigreeFunction	-0.03352267	0.13733730	0.04126495	0.18392757	0.18507093	0.14064695	1.00000000	0.03356131	0.17384407
Age	0.54434123	0.26351432	0.23952795	-0.11397026	-0.04216295	0.03624187	0.03356131	1.00000000	0.23835598
Outcome	0.22189815	0.46658140	0.06506836	0.07475223	0.13054795	0.29269466	0.17384407	0.23835598	1.00000000

Statistics:

Pregnancies - Glucose : 0.13
Pregnancies - BloodPressure : 0.14
Pregnancies - SkinThickness : -0.08
Pregnancies - Insulin : -0.07
Pregnancies - BMI : 0.02
Pregnancies - DiabetesPedigreeFunction : -0.03
Pregnancies - Age : 0.54
Pregnancies - Outcome : 0.22
Glucose - BloodPressure : 0.15
Glucose - SkinThickness : 0.06
Glucose - Insulin : 0.33
Glucose - BMI : 0.22
Glucose - DiabetesPedigreeFunction : 0.14
Glucose - Age : 0.26
Glucose - Outcome : 0.47
BloodPressure - SkinThickness : 0.21
BloodPressure - Insulin : 0.09
BloodPressure - BMI : 0.28
BloodPressure - DiabetesPedigreeFunction : 0.04
BloodPressure - Age : 0.24
BloodPressure - Outcome : 0.07
SkinThickness - Insulin : 0.44
SkinThickness - BMI : 0.39
SkinThickness - DiabetesPedigreeFunction : 0.18
SkinThickness - Age : -0.11
SkinThickness - Outcome : 0.07
Insulin - BMI : 0.2
Insulin - DiabetesPedigreeFunction : 0.19
Insulin - Age : -0.04
Insulin - Outcome : 0.13
BMI - DiabetesPedigreeFunction : 0.14
BMI - Age : 0.04
BMI - Outcome : 0.29
DiabetesPedigreeFunction - Age : 0.03
DiabetesPedigreeFunction - Outcome : 0.17
Age - Outcome : 0.24

Outcome - Pregnancies : 0.222
Outcome - Glucose : 0.467
Outcome - BloodPressure : 0.065
Outcome - SkinThickness : 0.075
Outcome - Insulin : 0.131
Outcome - BMI : 0.293
Outcome - DiabetesPedigreeFunction : 0.174
Outcome - Age : 0.238

Identified strong predictors (Glucose, BMI, DiabetesPedigreeFunction) vs. weak predictors (BloodPressure, SkinThickness) by comparing the correlation coefficient of all the variables with the outcome.

Glucose has the strongest positive correlation with diabetes Outcome.
BMI and Age have moderate correlation.
BloodPressure and SkinThickness are very weakly correlated with Outcome

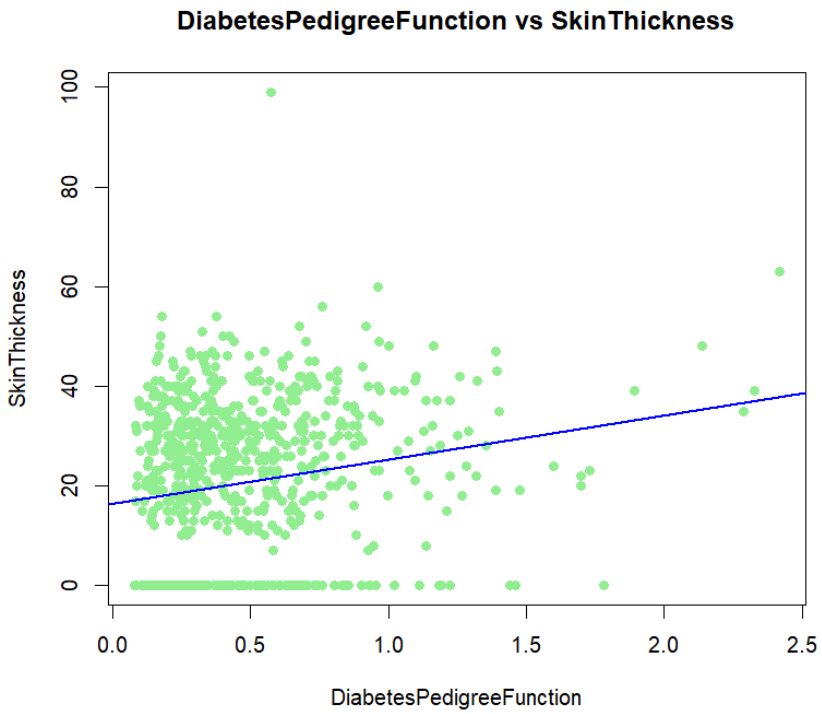
Sometimes a variable looks weak alone but can have interaction effects with stronger predictors. So plotting the following scatter plots and regression lines for validating whether the weak variables really matter in combination

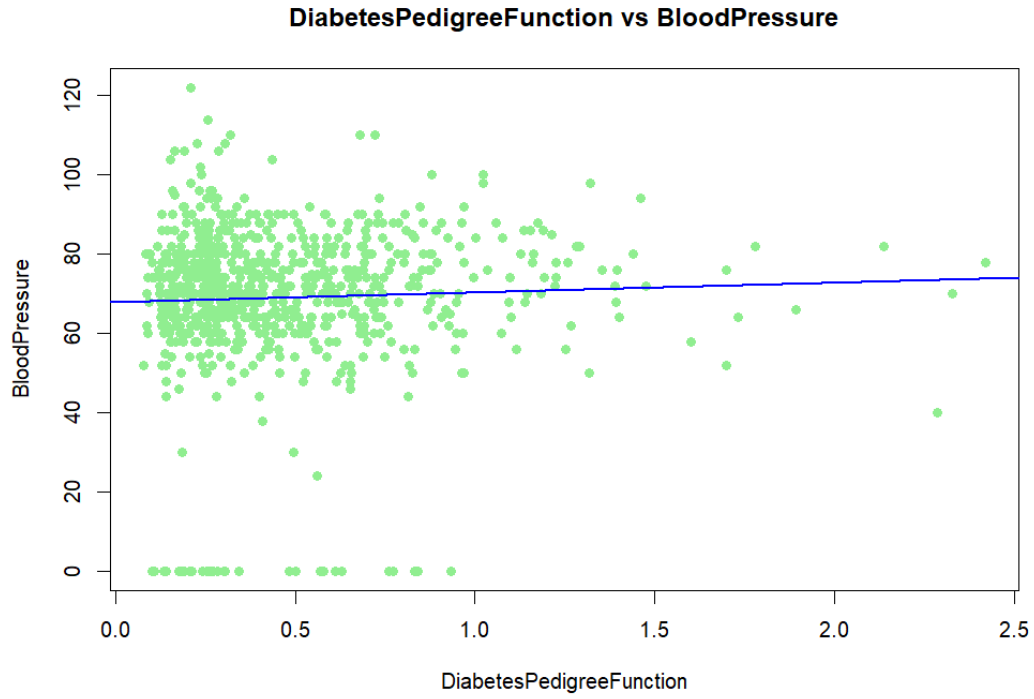
```
Glucose -> BloodPressure Slope: 0.092 R²: 0.023
Glucose -> SkinThickness Slope: 0.029 R²: 0.003
BMI -> BloodPressure Slope: 0.692 R²: 0.079
BMI -> SkinThickness Slope: 0.794 R²: 0.154
DiabetesPedigreeFunction -> BloodPressure Slope: 2.411 R²: 0.002
DiabetesPedigreeFunction -> SkinThickness Slope: 8.855 R²: 0.034
```

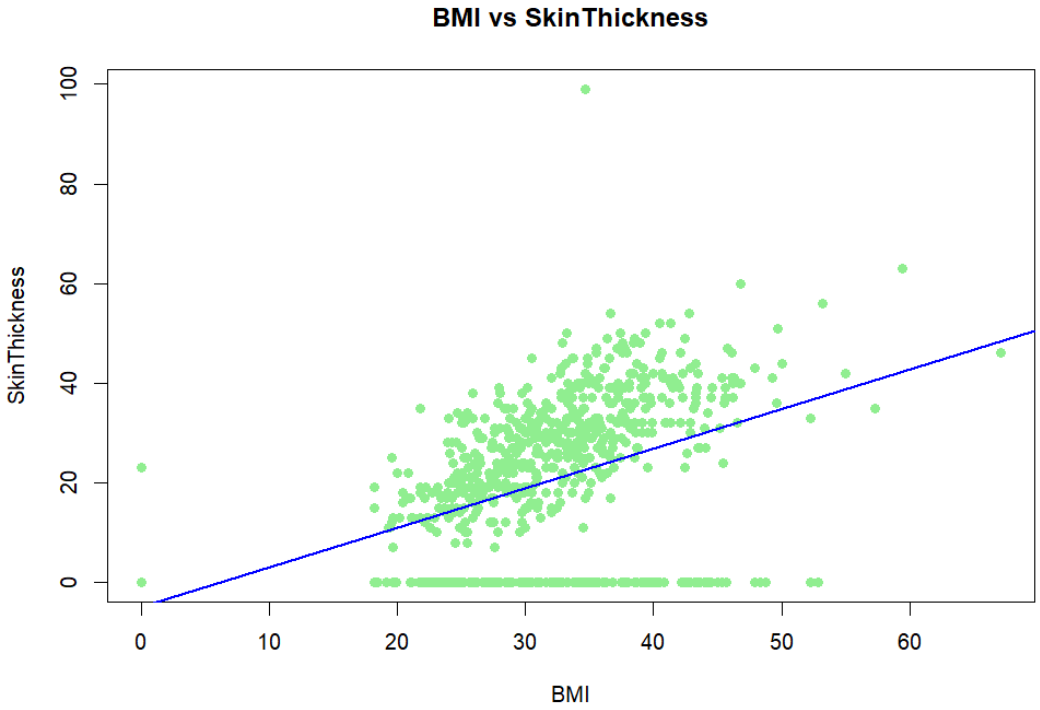
Most R^2 values are very low less than 0.2.

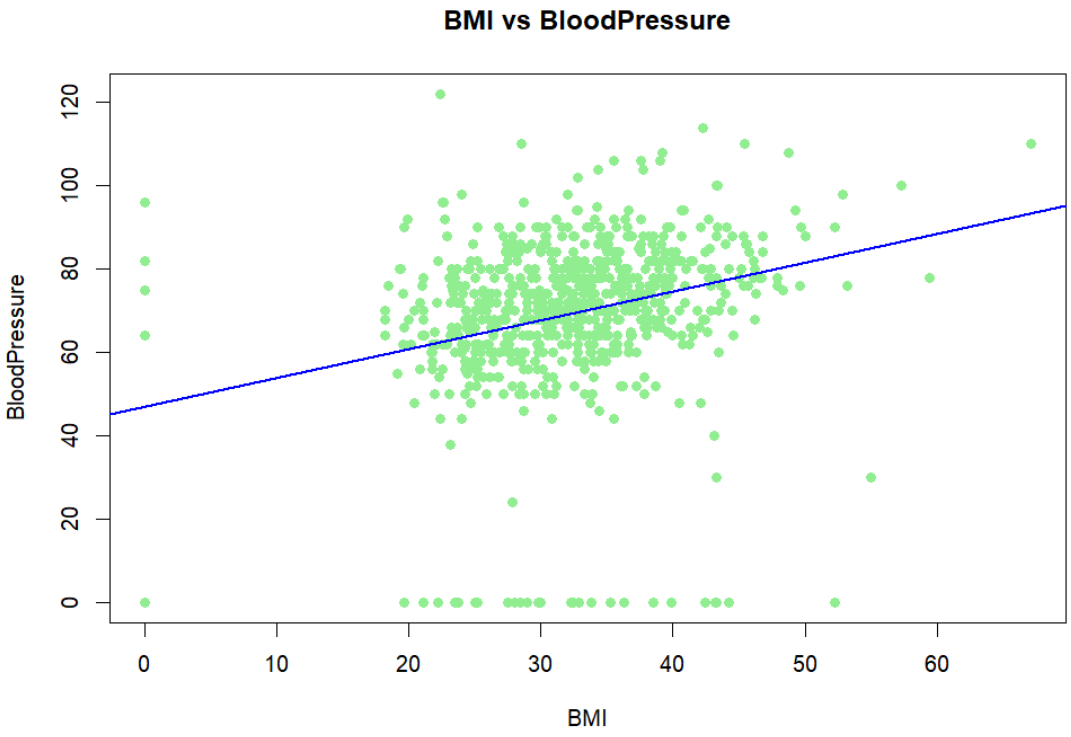
- This means weak predictors are not strongly dependent on strong predictors.
- They don't add much extra information via relationships with strong predictors.

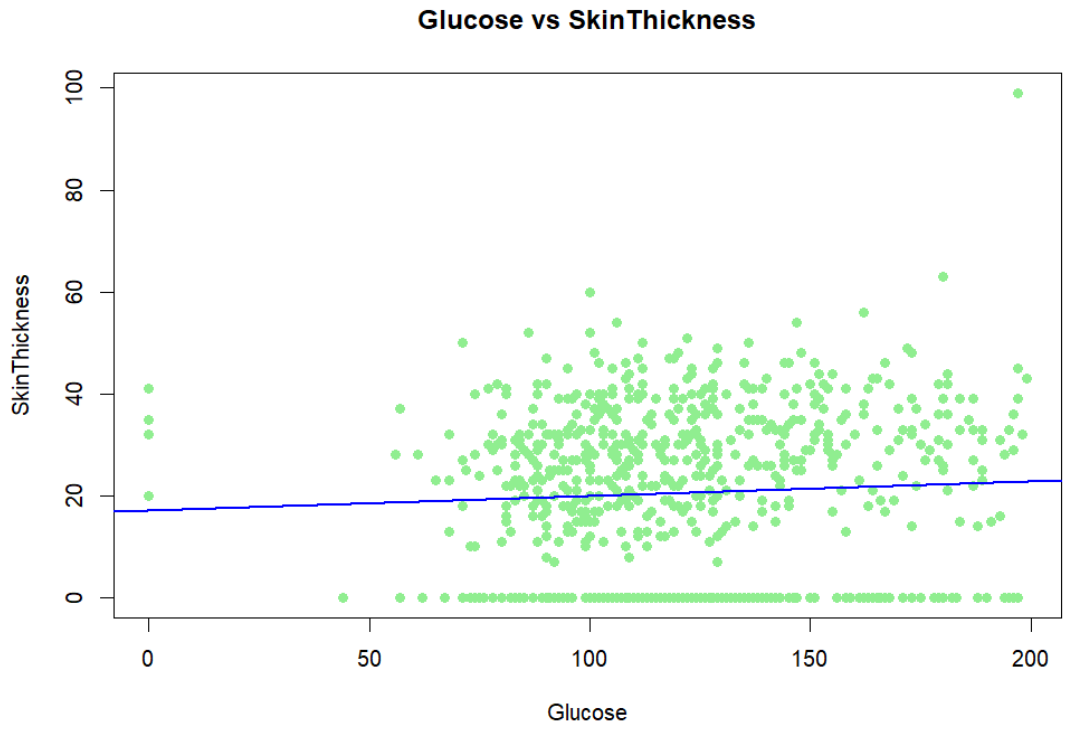
Plots:

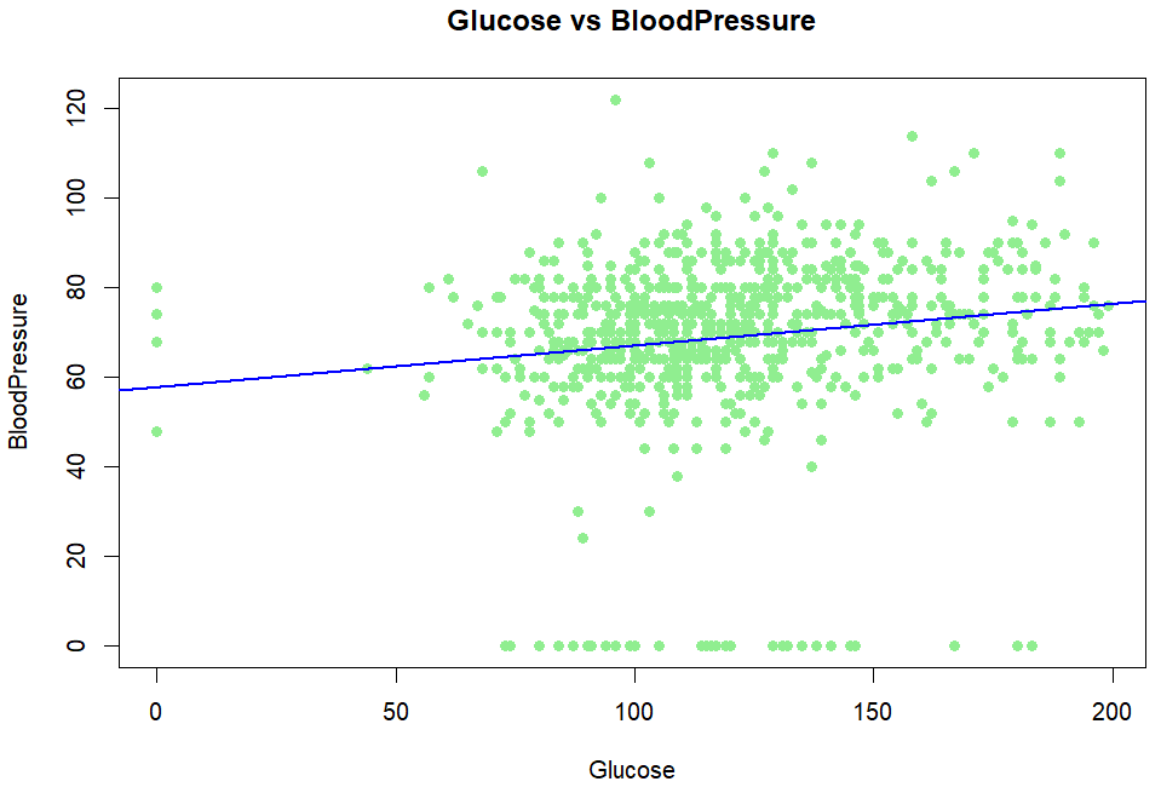



DiabetesPedigreeFunction vs BloodPressure
BloodPressure
0
20
40
60
80
100
120
0.0
0.5
1.0
1.5
2.0
2.5
DiabetesPedigreeFunction


BMI vs SkinThickness
SkinThickness
0
20
40
60
80
100
0
10
20
30
40
50
60
BMI


BMI vs BloodPressure
BloodPressure
0
20
40
60
80
100
120
0
10
20
30
40
50
60
BMI


Glucose vs SkinThickness
SkinThickness
0
20
40
60
80
100
0
50
100
150
200
Glucose



These relationships are important because sometimes weak predictors interact with strong ones in multivariate models.
Detecting these interactions can improve your predictive models later.
Even if a variable is weak alone, knowing its relationship with stronger variables helps you decide if it can be removed, kept, or transformed.
But in this data the weak variables don't contribute much in the outcome, neither directly nor indirectly.

- Glucose is the key predictor.
- BMI is moderately important.
- Weak predictors add very little explanatory power alone or in combination with strong predictors.