

Pima Indian Women: Diabetes Risk Analysis

Isabelle Berkowitz, Martin Topacio, Carly Schwartzberg, Siddharth Das

STA 135 Multivariate Data Analysis, June 6, 2024

Contents

1	INTRODUCTION	2
2	THE DATA	2
2.1	Description of Data Set	2
3	Initial Data Visualizations	2
3.1	Correlation Matrix Analysis	2
3.2	Box Plot Analysis	2
3.3	Box Plot Analysis	2
4	Methodology	4
5	Results and Discussion	4
5.1	Box's M-test for Homogeneity of Covariance Matrices	4
5.2	Quadratic Discriminant Analysis	4
5.3	Principal Component Analysis:	5
6	Conclusion	6
7	References	7

1 INTRODUCTION

In the 1960s the long-term commitment to study type 2 diabetes and obesity in Pima Indians began to indulge. With their willingness to participate in research exploration, statisticians identified significant findings regarding “the epidemiology, physiology, clinical assessment, and genetics”¹ heavily related to these major health issues. Throughout history, Pima Indians were forced to adapt to new territory due to white settlers, causing them to reset irrigation systems for agriculture to support their food growth. As a result of this displacement and farming issues, the amount of food, physical activity, and the height of their economy was very poor. As a result diabetes and obesity became more prevalent among the Pima Indians. The research collected showed that type 2 diabetes increased during this time for Pima women, likely as a result of increased body mass index. The goal of our project is to visually demonstrate the correlation between our dependent variable against various factors and to perform a more in-depth analysis of this data by applying multivariate techniques taught in the course and understanding why certain methods are or are not significant to our data.

2 THE DATA

In this project, we dive into the Pima Indians Diabetes Database to see what factors have higher significance in contributing to the prevalence of type 2 diabetes in female Pima Indians. The data has been collected from the U.S. National Institute of Diabetes and Digestive Kidney Diseases on women $21 \leq$. We extracted this data in R-Studio using the library package (**mlbench**) containing 768 observations onto 9 variables.² Computing the total counts in R, we found that **500** women do not have diabetes, while **268** do.

2.1 Description of Data Set

- Pregnancies: Number of times a woman has been pregnant
- Glucose: Plasma Glucose Concentration
- Pressure: Diastolic Blood Pressure (mm Hg)
- Triceps: Triceps Skin Fold Thickness (mm)
- Insulin: 2 Hour Serum Insulin (μ U/ml)
- Mass: Body mass index $\frac{weight(kg)}{height(m)^2}$
- Pedigree: Diabetes Pedigree Function
- Age: The age of the woman
- Diabetes: Represented as a factor level for the test results (2 = positive, 1 = negative)

3 Initial Data Visualizations

The following visualizations represent the beginning understanding of the relationships between diabetes prevalence among other factors.

3.1 Correlation Matrix Analysis

The first graphs demonstrate the correlation between factor variables and the distributions between the categorical variables, on whether a woman has diabetes or not. We see there is a high correlation between **age** and **pregnancy**, likely because as we get older, pregnancy becomes more common. There is some moderate correlation between **insulin** and **triceps**, implying that women with high insulin levels may have thicker triceps folds (and vice versa).

3.2 Box Plot Analysis

The second set of graphs represents the distribution of diabetes (pos/neg) of individuals among others. A significant finding is that young adults (**ages 20-30**) have a higher prevalence of diabetes-positive cases compared to aged $30 \leq$. **Higher glucose, BMI, and pedigree function values** associate with a **higher** incidence of diabetes. However, **blood pressure** and **tricep thickness** do **not** strongly correlate.

3.3 Box Plot Analysis

The boxplots represent the distribution of values for each factor on whether a person has diabetes(pos/neg), by identifying outliers. Showing **age** and **high glucose levels** are the most influential factors in diabetes prevalence over time. **Pregnancy** and **BMI** suggest significant differences among others, implying **lifestyle** heavily contributes to diabetes risk. In conclusion, these graphs confirm the well-known risk factors for diabetes outlining a clear comparison of these metrics between those with and without diabetes.

¹Dataset source: [High-Risk Populations: The Pimas of Arizona and Mexico](#)

²Dataset source: [Pima Indians Diabetes Data](#)

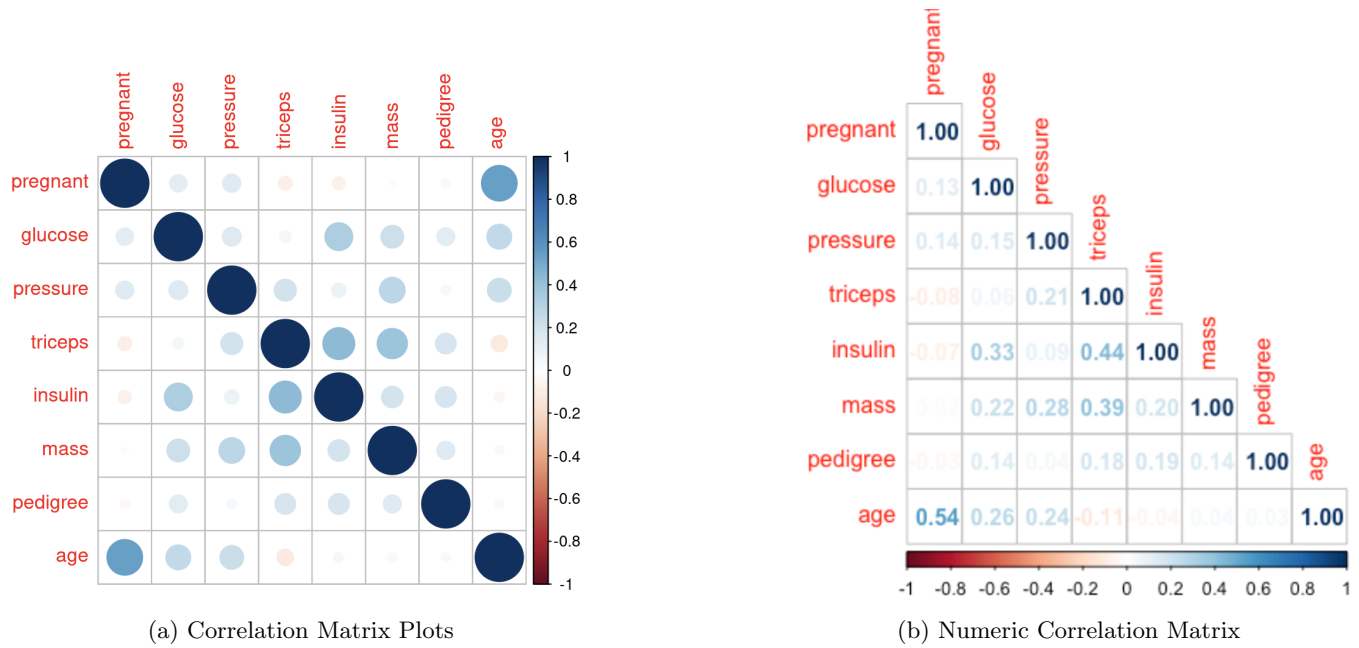


Figure 1: Correlation Matrix Plots to Understand Relationships between factors

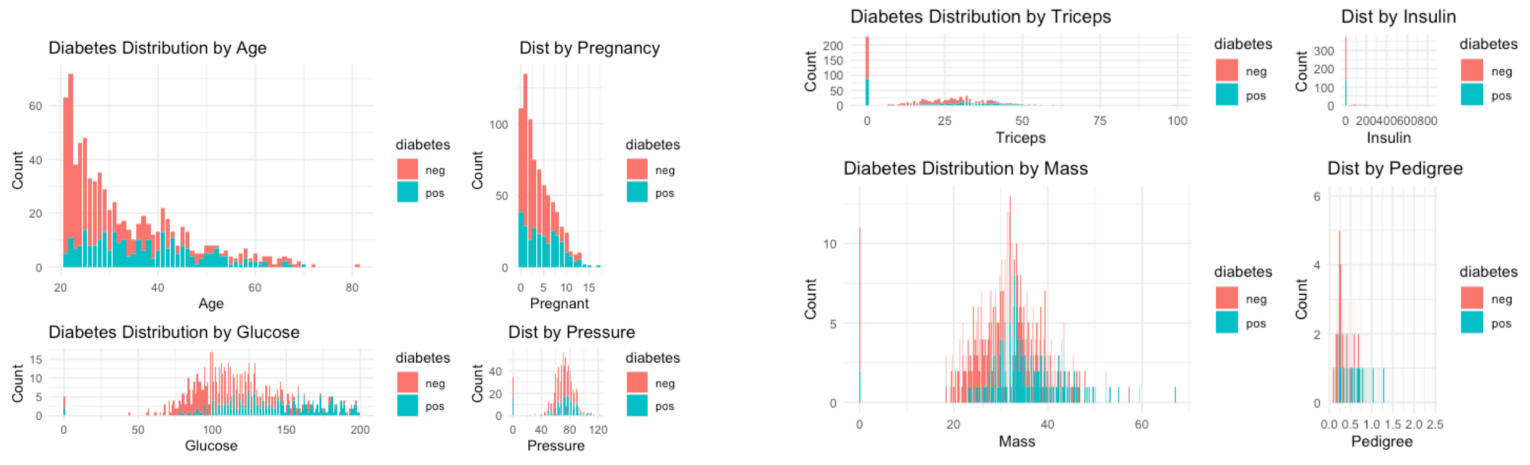


Figure 2: Distribution of first 4 Categorical variables

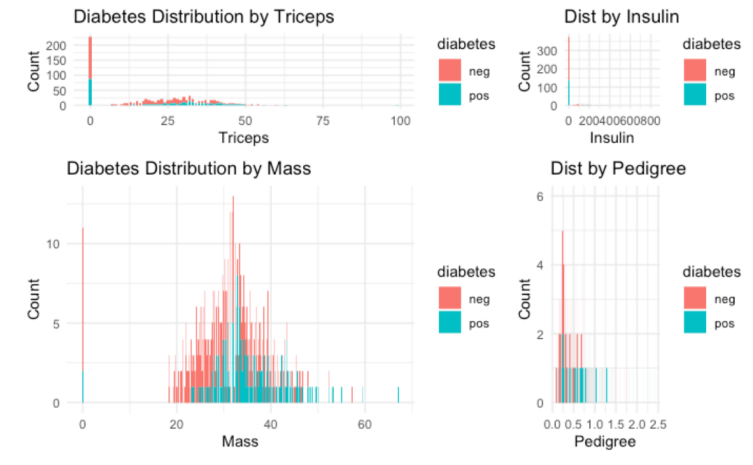
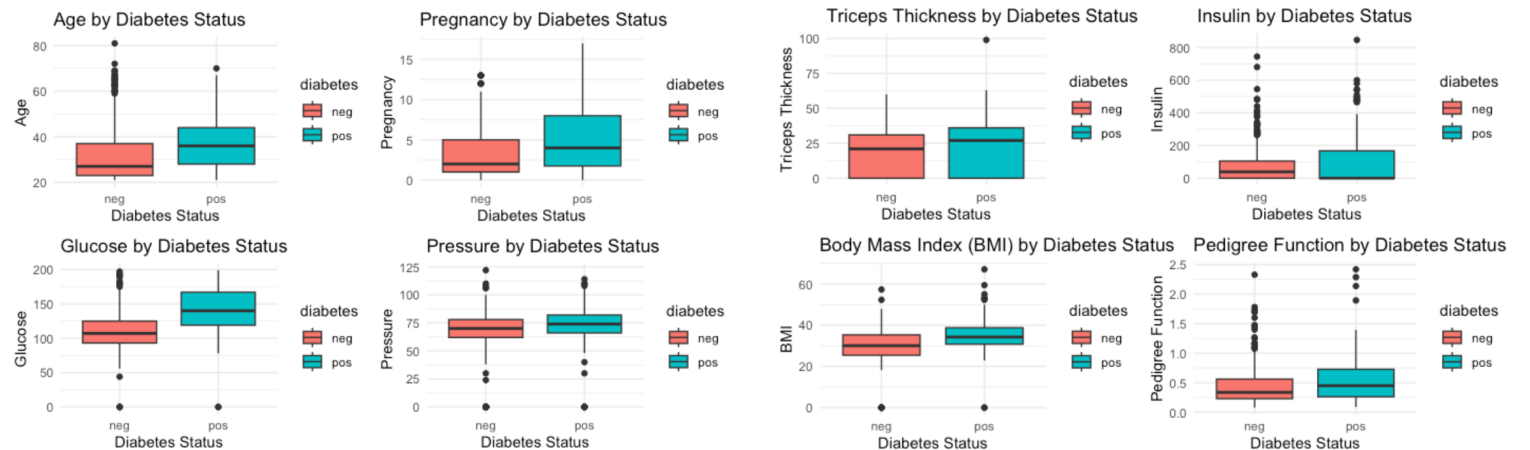


Figure 3: Distribution of last 4 Categorical variables



(a) Caption for Figure 1

(b) Caption for Figure 2

Figure 4: Overall Caption

4 Methodology

1. Box's M Test:

For the transitions of univariate to multivariate we conduct this test to check for the variance between the covariance matrices. This test can help statisticians determine if LDA or QDA applies better to specific data through hypothesis testing. When the **p-value** is < 0.05 we **reject the null hypothesis** stating that variance-covariance is the same. When the p-value is close to zero you will want to continue with QDA. This test assumes $x_1, x_2, \dots, x_{n1} \sim N_p(\mu_x, \Sigma_1)$ and $y_1, y_2, \dots, y_{n2} \sim N_p(\mu_y, \Sigma_2)$. These assumptions guide us into the computation of M:

$$M = \frac{|S_1|^{\frac{v_1}{2}} + |S_2|^{\frac{v_2}{2}}}{|S_{pl}|^{\frac{v_1}{2} + \frac{v_2}{2}}}$$

2. Linear Discriminant Analysis:

A is a form of statistical classification that identifies patterns and looks for 'linear combinations that separate different factors between classes and data types.'³ Our model used what is called **fault diagnosis** as we attempted to see which variables were **"good"** = high association and **"bad"** = low effect in diabetes prevalence. Depending on the outcome of the p-value in Box's M test we determine if LDA or QDA is more reasonable. To check assumptions for this model, you can implement Box's M test along with the linear discriminant function:

$$s_k(x) = x^T \Sigma^{-1} \mu_k - 0.5 \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

3. Quadratic Discriminant Analysis

On the other hand, QDA allows us to understand how our specified factors vary when the population covariance matrices are different. In the form of classification, this procedure leads to conclusions as to how variables discriminate between groups of dependent and controlled data. These results tend to be more accurate and credible which is significant in looking at risk factors such as diabetes prevalence among risk and demographic factors. The quadratic discriminant function follows as:

$$S_k(x) = \frac{-1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log \pi_k$$

4. Principal Component Analysis

PCA is a common statistical procedure requiring us to "choose a subspace to maximize the projected variance, or minimize the reconstruction error."⁴ Once we remove constants to avoid misleading results, the test can be performed. We are then left with standard deviations on the principal components, variance, and scaled proportions.⁵ PCA applies the **Courant-Fischer Theorem** where if **A** is a symmetric matrix, we denote the eigenvalues as $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ where:

$$\lambda_k \min_{\alpha \in T_k} \cdot \max_{x \in S_k^\alpha} \cdot k(\text{zerovector}) \frac{X^T A x}{\|x\|_2^2}$$

5 Results and Discussion

5.1 Box's M-test for Homogeneity of Covariance Matrices

We use this to determine whether we should implement QDA or LDA. When we conducted this test in R we got a p-value of $< 2.2\text{e-}16$, which is so close to zero that we rejected the null hypothesis and continued with performing QDA on the dataset.

5.2 Quadratic Discriminant Analysis

By implementing QDA we split the data into training and test sets. This allowed us to perform QDA with k-fold cross-validation, and again without splitting data. We plan to evaluate the accuracy of each QDA process and see which is the best for fitting the data, new and old. From the confusion matrix, we see that it yields **122** true negatives, **28** false negatives, **41** true positives, and **39** false positives. False positives are indicative of Type I error, while false negatives are indicative of Type II error. From our confusion matrix, we see **39** healthy women were diagnosed with onset diabetes, and **28** women with onset diabetes were incorrectly diagnosed as healthy.

³Dataset source: [How to Use LDA and QDA with Multiple Predictors](#)

⁴Dataset source: [Week 8-1.annotated](#)

⁵Dataset source: [What is Principal Component Analysis](#)

a) Confusion Matrix:

From the confusion matrix, we see that it yielded **122** true negatives and **28** false negatives, **41** true positives, and **39** false positives. False positives are indicative of Type I error, and false negatives are indicative of Type II error. This means from our confusion matrix, **39** healthy women were diagnosed with onset diabetes, and 28 women with onset diabetes were not.

b) QDA (split) Performance Metrics:

In this table (outputted in markdown) we identify several key findings. First, we obtain an accuracy rate of **70.87%**, which is the %age of correctly classified instances. We get a precision rate of **75.78%**, which is the %age of identifications that prove to be correct. The recall %age is **81.33%**, the %age of actual positives that are correctly identified. The F1 score, the mean of precision and recall, is 78.46%. Finally, we obtained a misclassification error rate of **29.13%**. For precision, recall, and F1 score, we ideally would desire rates at 80 % or above. Our precision and F1 rates are somewhat close to 80 %, while our recall %age is over 80 %, which is generally good. We want high accuracy, ideally above **90%**, but our accuracy is at **70.87%**, which is not good enough. Our misclassification error rate is mediocre at **29.13%**, with a rate at **40%** or above being considered poor.

c) QDA with cross-validation - Confusion Matrix

Examining the confusion matrix from the QDA model with cross-validation, we have **432** true negatives **68** false negatives, **155** true positives and **113** false positives. This indicates that **113** healthy women were incorrectly diagnosed with diabetes (Type I error), and **68** women with diabetes were incorrectly diagnosed as healthy (Type II error).

d) QDA with cross-validation - Performance Metrics

When running a QDA model with cross-validation, we obtain several findings. First, our accuracy is **76.43%**, showing improvement over the prior model **WITH** training and test sets. There are improved precision and recall rates at **79.27%** and **86.4%**, respectively, an improved F1 score of **82.68%**, and a lower misclassification error rate, **23.57%**. Our accuracy rate is about **6%** closer to a "good" rate of 80 %, but not at our desired "great" rate of 90% or higher. Our precision rate is almost at 80 %, considered generally good, and our recall rate is sitting at a good % age. Our F1 score is also now above 80 %, which is good. The misclassification error rate did not drop as much as we hoped, with it still sitting at a somewhat mediocre %age, but it is still beneficial that it lowered at all.

e) QDA without splitting - Performance Metrics

Since both the QDA cross-validation model and the QDA model without splitting result in the same accuracy, precision, recall, F1 score, and misclassification error rate, the cross-validation model is better because it has less risk of overfitting, so it will make more accurate predictions on new data compared to the model without splitting data.

5.3 Principal Component Analysis:

We implemented PCA in our project which allowed us to explore relationships in a specific subspace to optimize projection variance. Our goal of PCA was to minimize the reconstruction error.

a) PCA Plots (PC1 vs. PC2): Plot 1 shows the correlation between PC1 and PC2. From this(in markdown), we know PC1 is heavily driven by insulin levels and PC2 is driven by glucose. The cluster at the origin suggests, that most people have average levels of both. Those straying right on the x-axis indicate higher insulin and those straying up on the y-axis indicate higher glucose levels. From this plot, we conclude the individuals who are at risk for both factors are the most likely to have the risk of diabetes.

b) Contribution Ratio: The contribution ratio plot shows us how many principal components are needed to get a majority of the variance. We can see a fast incline from the first 2 points and relative steadiness among other components, not to much fluctuation. Showing us that PC1 and PC2 are the main factors for the variance. Overall from this plot, we can see that after the first 6 components, the plot flattens out.

c) Overall Threshold: For the overall threshold, we interpret the eigenvalues and see there is a large drop from PC1 to PC2. The line represents the mean and those above it have significance to the variance. The graph shows only PC1 is above the line, and in far second place PC2 is closest to the line out of the remaining components.

d)Scree Plot: The scree plot shows us a very similar thing to the overall threshold plot. Instead of looking at the big drop in eigenvalues to interpret the number of components needed, the Scree plot gives us a bigger overview of the eigenvalues for our dataset. This shows us the rate of variation as you go from one component to the next. As we can see here the output is generally the same showing the eigenvalue for PC1 to be by far the highest.

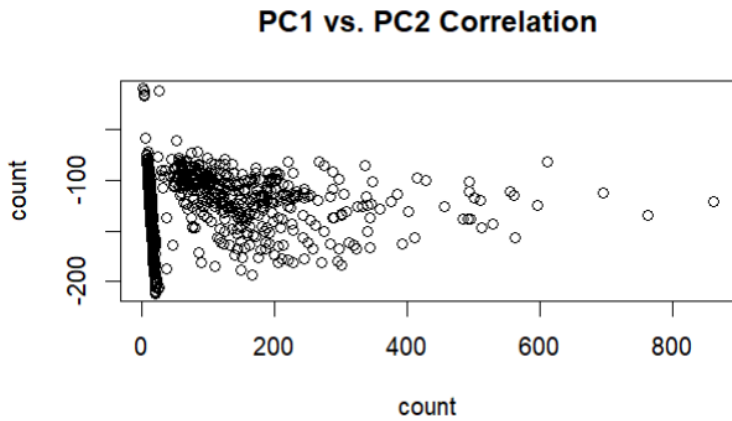


Figure 5: a) PC1, PC2 Correlation

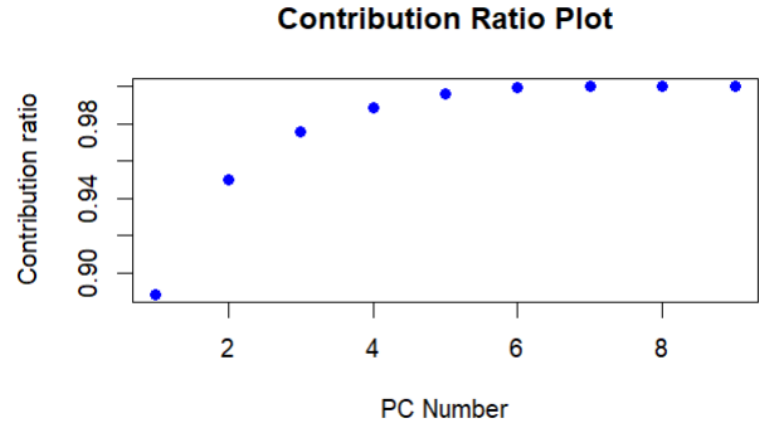


Figure 6: b) PC Relation to Variance

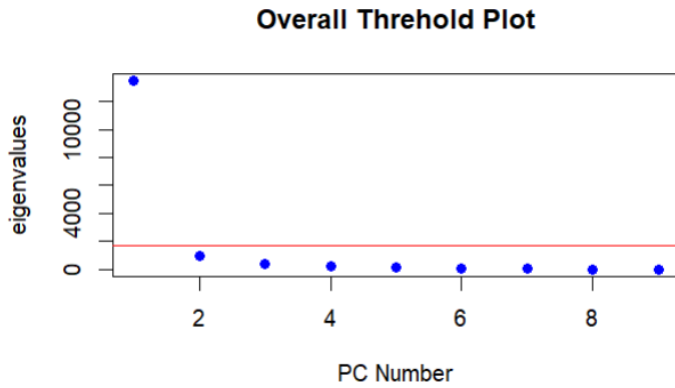


Figure 7: c) Comparison of Trends in Eigenvalues

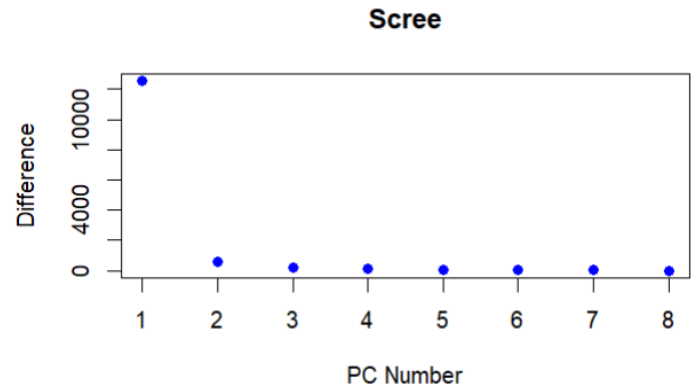


Figure 8: d) Overview of the Eigenvalues

6 Conclusion

In the 1960s the long term-commitment to study type 2 diabetes and obesity in women Pima Indians began to indulge. With their homes taken over by American settlers throughout history, their access to food, levels of physical activity, and economy was poor. Our project visually demonstrated the correlation between our dependent variable (diabetes) against our various factors and showed a more in-depth analysis on this data by applying multivariate techniques learned in the course and understanding why certain methods are or are not significant to our data. We accessed our data using the library package (mlbench) which containing data on 768 women with 9 observations. Our initial exploratory data analysis showed us that correlations exist between age and number of pregnancies, and insulin levels and triceps fold thickness. Furthermore, we noticed an increased occurrence of diabetes in young adults between 20 and 30 years old. In terms of multivariate methodology, we utilized the Box's M Test to learn whether the covariance matrices are the same. Since the covariance matrices were different, we developed three QDA models and assessed their performance metrics. Finally, we implemented PCA and found that insulin and glucose, or the first 2 principal components, explained a majority of the variance in the data. One key finding from our QDA models was that regardless of k-fold cross validation, our predictions were almost identical in terms of accuracy. Despite that, we prefer the k-fold cross validation model because it generally performs better while accounting for overfitting and evaluation bias. All in all, this study is beneficial for understanding the risk factors for type 2 diabetes among Pima Indian Women. This knowledge can benefit society by guiding health policies and developing more effective diabetes-prevention strategies. In the future, we intend to further research how health policies can be improved to be more inclusive of minority groups. Isabelle B produced and formatted the report in LaTeX by writing the introduction, data summaries, initial data visualizations, and methodology. In addition she wrote the code that produces all of our initial graphs with an in-depth analysis on those results.

CONTRIBUTIONS: Carly and Martin worked on conducting our multivariate procedures by writing the code and presenting those results under the results in discussion section. They both made additional graphs that are included in our markdown based on their findings. Finally, Sid worked help put the report together by tying all of our work together in the conclusion and performing test on QDA with Martin. Note: We utilized Professor Ding's Lecture Notes to help with the methodology, as well as Chat GPT to help with our testing and visualizations.

7 References

- ChatGPT (For assistance in modeling graphs/errors in LaTeX/some Wording, June 4, 2024).
- Ding Xiuxai (2024) Week 7-1 [HTML Lecture Slides] University of California, Davis.
<https://canvas.ucdavis.edu/courses/877214/files/folder/Lecture%20notes?preview=24340030>. Accessed 6 June 2024.
- Ding Xiuxai (2024) Week 8-1 [HTML Lecture Slides] University of California, Davis.
<https://canvas.ucdavis.edu/courses/877214/files/folder/Lecture%20notes?preview=24492190>. Accessed 6 June 2024.
- “Pima Indians Diabetes Data.” R, search.r-project.org/CRAN/refmans/pdp/html/pima.html. Accessed 11 June 2024.
- Renzo, Charles Di. “How to Use LDA and QDA with Multiple Predictors.” LinkedIn, 29 Feb. 2024, www.linkedin.com/pulse/how-use-lda-qda-multiple-predictors-charles-di-renzo-djrgc/.
- Schulz, Leslie O, and Lisa S Chaudhari. “High-Risk Populations: The Pimas of Arizona and Mexico.” Current Obesity Reports, U.S. National Library of Medicine, Mar. 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4418458/. Accessed 8 June, 2024
- “What Is Principal Component Analysis (PCA) and How It Is Used?” Sartorius, 8 Aug. 2020, www.sartorius.com/en/knowledge/science/snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186. Accessed 7 June 2024.

STA 135 Final Project

Martin Topacio, Isabelle Berkowitz, Carly Schwartzberg, Siddharth Das

2024-06-05

```
library(mlbench)
library(MASS)
library(caret)

## Loading required package: ggplot2
## Loading required package: lattice
library(corrplot)

## corrplot 0.92 loaded
library(randomForest)

## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:ggplot2':
##
##     margin
library(biotools)

## ---
## biotools version 4.2
library(e1071)

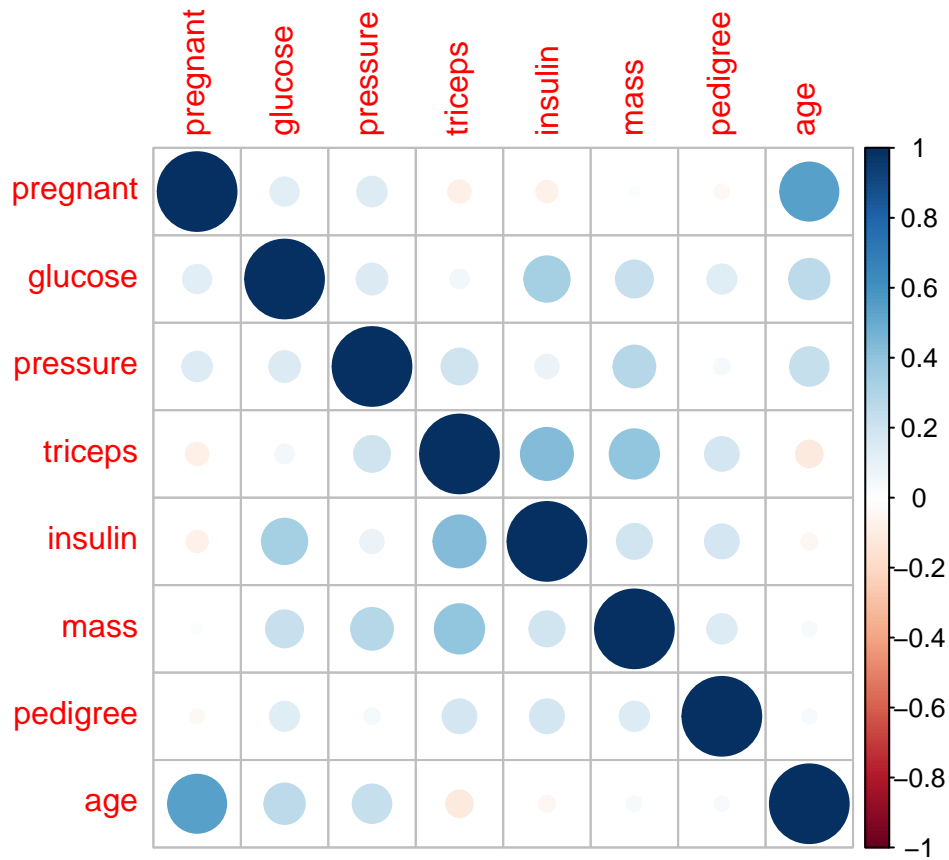
data(PimaIndiansDiabetes)
head(PimaIndiansDiabetes)

##   pregnant glucose pressure triceps insulin mass pedigree age diabetes
## 1         6     148      72      35        0 33.6    0.627  50      pos
## 2         1      85      66      29        0 26.6    0.351  31      neg
## 3         8     183      64       0        0 23.3    0.672  32      pos
## 4         1      89      66      23       94 28.1    0.167  21      neg
## 5         0     137      40      35      168 43.1    2.288  33      pos
## 6         5     116      74       0        0 25.6    0.201  30      neg
dim(PimaIndiansDiabetes)

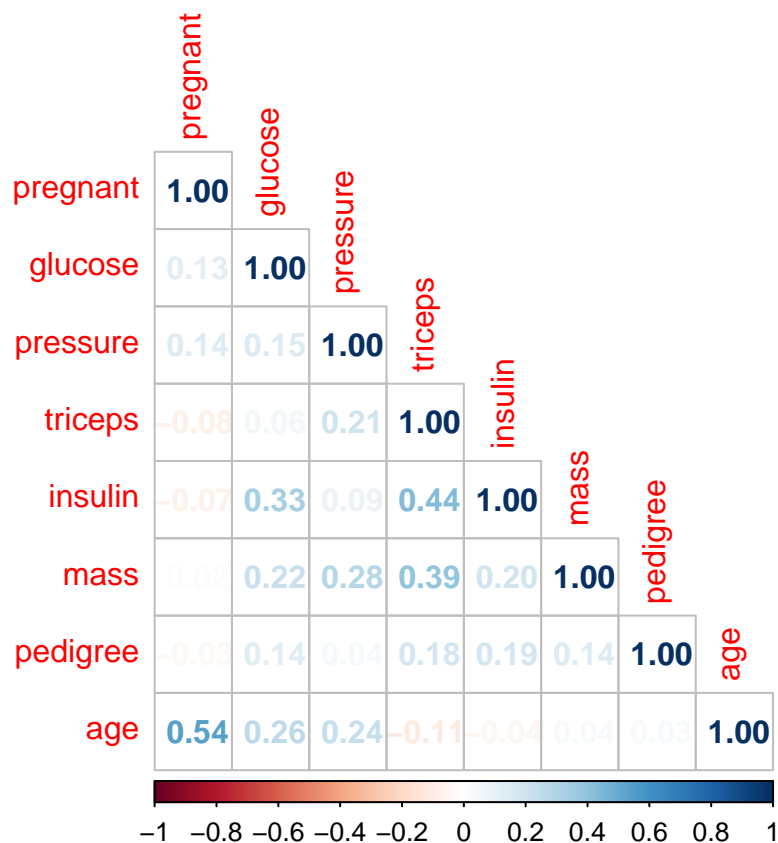
## [1] 768  9
```

Visualizations


```
library(corrplot)
correlation_matrix <- cor(PimaIndiansDiabetes[, 1:8], use = "complete.obs")
corrplot(correlation_matrix, method = "circle")
```



```
corrplot::corrplot(correlation_matrix, type = "lower", method = "number")
```



```
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:randomForest':
##
##      combine

plot_age <- ggplot(PimaIndiansDiabetes, aes(x = age, fill = diabetes)) +
  geom_bar() +
  labs(title = "Diabetes Distribution by Age", x = "Age", y = "Count") +
  theme_minimal()

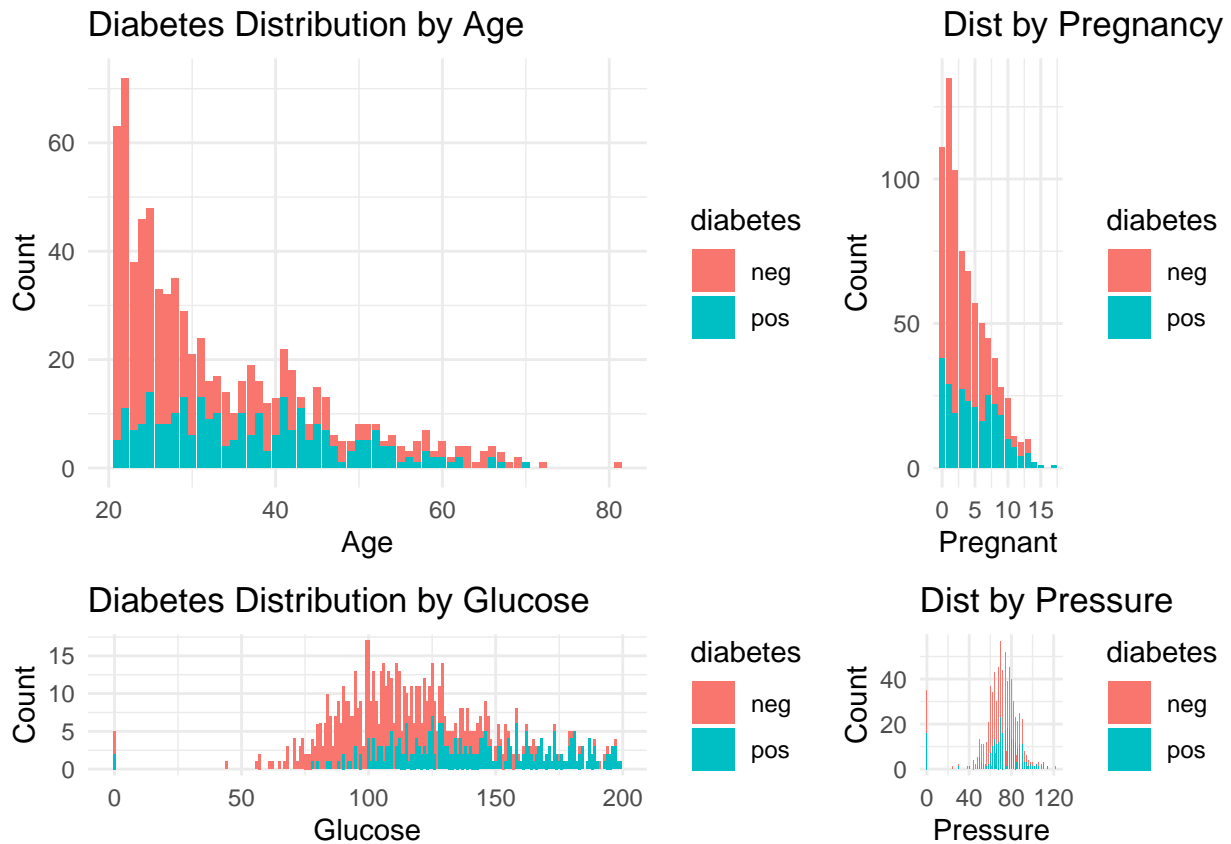
plot_pregnant <- ggplot(PimaIndiansDiabetes, aes(x = pregnant, fill = diabetes)) +
  geom_bar() +
  labs(title = "Dist by Pregnancy", x = "Pregnant", y = "Count") +
  theme_minimal()

plot_glucose <- ggplot(PimaIndiansDiabetes, aes(x = glucose, fill = diabetes)) +
  geom_bar() +
  labs(title = "Diabetes Distribution by Glucose", x = "Glucose", y = "Count") +
  theme_minimal()

plot_pressure <- ggplot(PimaIndiansDiabetes, aes(x = pressure, fill = diabetes)) +
  geom_bar() +
  labs(title = "Dist by Pressure", x = "Pressure", y = "Count") +
```

```
theme_minimal()

#space out graphs so not in one chunk
grid.arrange(plot_age, plot_pregnant, plot_glucose, plot_pressure,
              ncol = 2, nrow = 2,
              heights = c(2, 1), widths = c(2, 1))
```



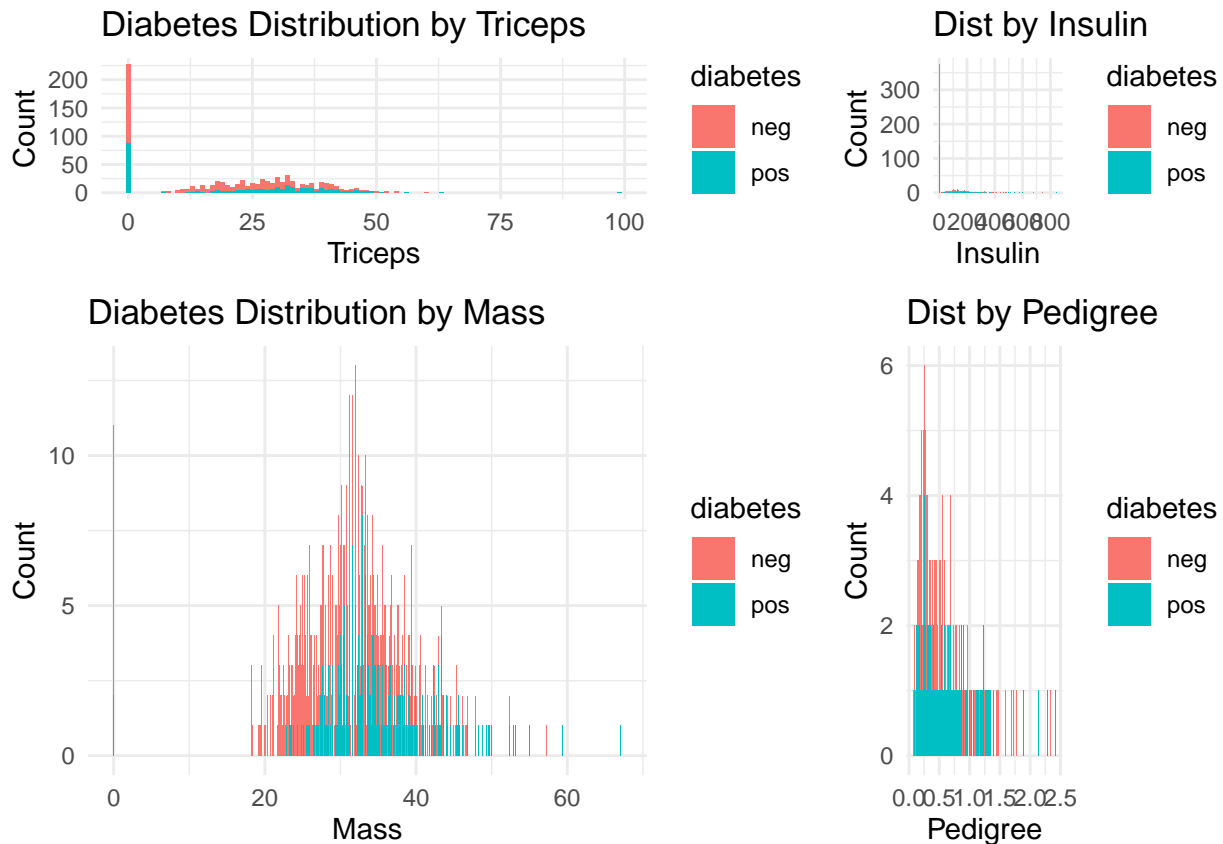
```
plot_triceps <- ggplot(PimaIndiansDiabetes, aes(x = triceps, fill = diabetes)) +
  geom_bar() +
  labs(title = "Diabetes Distribution by Triceps", x = "Triceps", y = "Count") +
  theme_minimal()

plot_insulin <- ggplot(PimaIndiansDiabetes, aes(x = insulin, fill = diabetes)) +
  geom_bar() +
  labs(title = "Dist by Insulin", x = "Insulin", y = "Count") +
  theme_minimal()

plot_mass <- ggplot(PimaIndiansDiabetes, aes(x = mass, fill = diabetes)) +
  geom_bar() +
  labs(title = "Diabetes Distribution by Mass", x = "Mass", y = "Count") +
  theme_minimal()

plot_pedigree <- ggplot(PimaIndiansDiabetes, aes(x = pedigree, fill = diabetes)) +
  geom_bar() +
  labs(title = "Dist by Pedigree", x = "Pedigree", y = "Count") +
  theme_minimal()
```

```
#space out plots better
grid.arrange(plot_triceps, plot_insulin, plot_mass, plot_pedigree,
             ncol = 2, nrow = 2,
             heights = c(1, 2), widths = c(2, 1))
```



```
library(ggplot2)
plot1 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = age, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Age by Diabetes Status", x = "Diabetes Status", y = "Age") +
  theme_minimal()

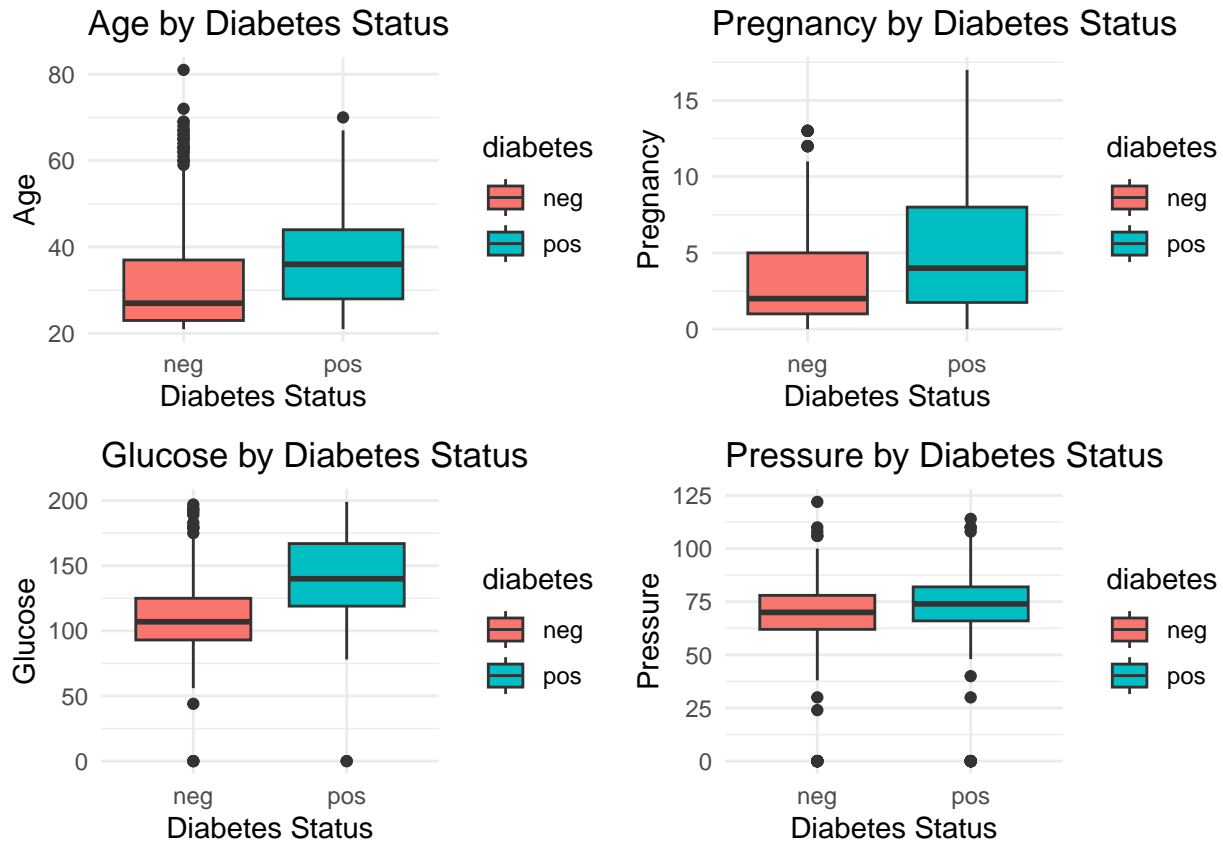
plot2 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = pregnant, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Pregnancy by Diabetes Status", x = "Diabetes Status", y = "Pregnancy") +
  theme_minimal()

plot3 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = glucose, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Glucose by Diabetes Status", x = "Diabetes Status", y = "Glucose") +
  theme_minimal()

plot4 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = pressure, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Pressure by Diabetes Status", x = "Diabetes Status", y = "Pressure") +
  theme_minimal()

# align plots together
library(gridExtra)
```

```
grid.arrange(plot1, plot2, plot3, plot4, ncol = 2)
```



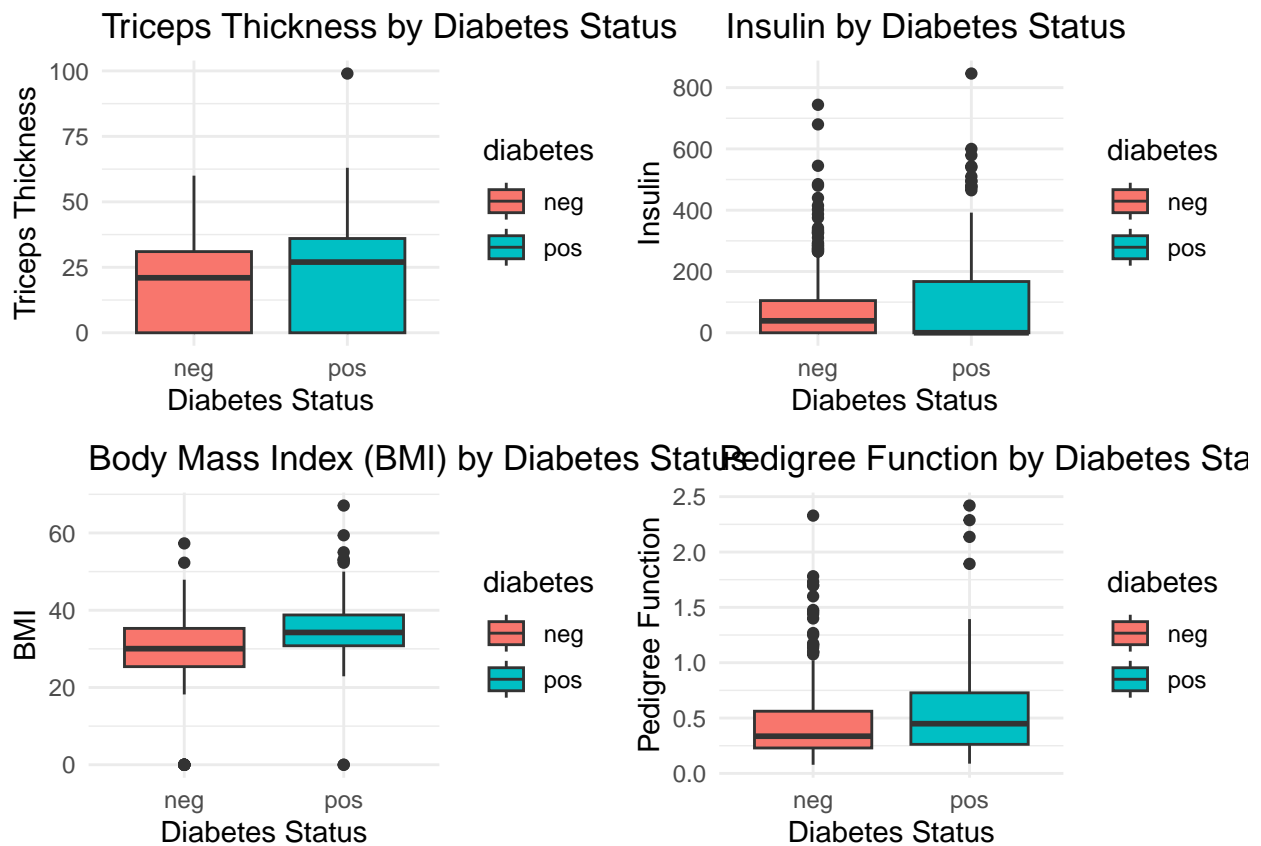
```
plot5 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = triceps, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Triceps Thickness by Diabetes Status", x = "Diabetes Status", y = "Triceps Thickness") +
  theme_minimal()

plot6 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = insulin, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Insulin by Diabetes Status", x = "Diabetes Status", y = "Insulin") +
  theme_minimal()

plot7 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = mass, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Body Mass Index (BMI) by Diabetes Status", x = "Diabetes Status", y = "BMI") +
  theme_minimal()

plot8 <- ggplot(PimaIndiansDiabetes, aes(x = diabetes, y = pedigree, fill = diabetes)) +
  geom_boxplot() +
  labs(title = "Pedigree Function by Diabetes Status", x = "Diabetes Status", y = "Pedigree Function") +
  theme_minimal()

# align plots side by side
grid.arrange(plot5, plot6, plot7, plot8, ncol = 2)
```



Box's M Test

```
# Load necessary libraries
library(mlbench)
library(MASS)
library(caret)
library(biotools) # for boxM function

# Load the dataset
data(PimaIndiansDiabetes)

# Check the structure of the dataset and the unique values in the `diabetes` column
str(PimaIndiansDiabetes)
```

```
## 'data.frame': 768 obs. of 9 variables:
## $ pregnant: num 6 1 8 1 0 5 3 10 2 8 ...
## $ glucose : num 148 85 183 89 137 116 78 115 197 125 ...
## $ pressure: num 72 66 64 66 40 74 50 0 70 96 ...
## $ triceps : num 35 29 0 23 35 0 32 0 45 0 ...
## $ insulin : num 0 0 0 94 168 0 88 0 543 0 ...
## $ mass : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
## $ pedigree: num 0.627 0.351 0.672 0.167 2.288 ...
## $ age : num 50 31 32 21 33 30 26 29 53 54 ...
## $ diabetes: Factor w/ 2 levels "neg","pos": 2 1 2 1 2 1 2 1 2 2 ...

print(unique(PimaIndiansDiabetes$diabetes))
```

```

## [1] pos neg
## Levels: neg pos

# Ensure the outcome variable is a factor
PimaIndiansDiabetes$diabetes <- as.factor(PimaIndiansDiabetes$diabetes)

# Ensure the data is numeric (including the outcome variable)
PimaIndiansDiabetes <- PimaIndiansDiabetes[, sapply(PimaIndiansDiabetes, is.numeric) | colnames(PimaIndiansDiabetes) == "diabetes"]

# Split the data into two groups based on diabetes status
group1 <- PimaIndiansDiabetes[PimaIndiansDiabetes$diabetes == "pos", ]
group2 <- PimaIndiansDiabetes[PimaIndiansDiabetes$diabetes == "neg", ]

# Check the number of rows in each group to ensure they are not empty
print(paste("Number of rows in group1 (diabetes = pos):", nrow(group1)))

## [1] "Number of rows in group1 (diabetes = pos): 268"

print(paste("Number of rows in group2 (diabetes = neg):", nrow(group2)))

## [1] "Number of rows in group2 (diabetes = neg): 500"

# Check for near-zero variance predictors and remove them
nzv <- nearZeroVar(PimaIndiansDiabetes, saveMetrics = TRUE)
nzv_cols <- rownames(nzv[nzv$nzv == TRUE, ])
if (length(nzv_cols) > 0) {
  group1 <- group1[, !(colnames(group1) %in% nzv_cols)]
  group2 <- group2[, !(colnames(group2) %in% nzv_cols)]
}

# Ensure there are no issues with the subsetted groups
if (nrow(group1) == 0 || nrow(group2) == 0) {
  stop("One of the groups is empty after subsetting.")
}

# Remove the outcome variable from both groups
group1 <- group1[, -which(names(group1) == "diabetes")]
group2 <- group2[, -which(names(group2) == "diabetes")]

# Combine the data into a single data frame and create a grouping factor
combined_data <- rbind(group1, group2)
grouping <- factor(c(rep("pos", nrow(group1)), rep("neg", nrow(group2))))

# Perform Box's M test
result <- boxM(combined_data, grouping)

# Print the result
print(result)

##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data: combined_data
## Chi-Sq (approx.) = 226.71, df = 36, p-value < 2.2e-16

```

Since we obtain a p-value so close to zero, we reject the null hypothesis and continue on with performing QDA on the data.

QDA:

For the QDA process, we aim to first perform QDA by simply splitting the data into training and test sets, then we will perform QDA with k-fold cross-validation, then we will perform QDA without any splitting. We plan to evaluate the accuracy of each QDA process and see which one works best in terms of fitting the data, new and old.

QDA - splitting into training and test sets

```
cov(group1)
```

```
##           pregnant      glucose      pressure      triceps      insulin
## pregnant  13.99686958   -6.5233384   10.2085639   -5.236290   -40.76371
## glucose   -6.52333836  1020.1394572   47.1576807   21.242216  1157.77464
## pressure  10.20856392   47.1576807  461.8979680   85.586953   266.37372
## triceps   -5.23629046   21.2422159  85.5869529   312.572195  1119.47275
## insulin   -40.76370954  1157.7746380  266.3737213  1119.472749 19234.67332
## mass      -4.32235452   11.6957488   20.9089860   40.070518   55.51413
## pedigree  -0.09639326   0.3148521    0.2762678    1.803116    5.24497
## age       18.25999217   34.5294594   61.9219353   -17.842529   36.42305
##           mass      pedigree      age
## pregnant  -4.3223545 -0.09639326  18.2599922
## glucose    11.6957488  0.31485206  34.5294594
## pressure   20.9089860  0.27626779  61.9219353
## triceps    40.0705182  1.80311610 -17.8425289
## insulin    55.5141260  5.24497004  36.4230533
## mass       52.7506932  0.36985506 -14.9773995
## pedigree    0.3698551  0.13864786 -0.3598839
## age       -14.9773995 -0.35988390 120.3025882
```

```
cov(group2)
```

```
##           pregnant      glucose      pressure      triceps      insulin
## pregnant  9.10340281   7.7835271   7.2536754   -5.3165050   -39.370758
## glucose   7.78352705  683.3623246  91.0357515   6.2337475   912.202244
## pressure  7.25367535  91.0357515  326.2746934  50.3144529  133.268810
## triceps   -5.31650501  6.2337475  50.3144529  221.7105251  607.667447
## insulin   -39.37075752  912.2022445  133.2688096  607.6674469  9774.345427
## mass       0.38271383  26.4844529  50.4463198  50.2210533  193.259192
## pedigree  -0.07214703  0.7470368   0.1474398   0.4238764   6.723592
## age       20.16370741  69.5468938  45.2475351  -28.4550701 -172.144770
##           mass      pedigree      age
## pregnant  0.3827138 -0.07214703  20.1637074
## glucose   26.4844529  0.74703675  69.5468938
## pressure  50.4463198  0.14743982  45.2475351
## triceps   50.2210533  0.42387638  -28.4550701
## insulin   193.2591920  6.72359186 -172.1447695
## mass      59.1338701  0.16252256   3.2362745
## pedigree  0.1625226  0.08945202   0.1453953
## age      3.2362745  0.14539533  136.1341683
```

```
colMeans(group1)
```

```
##      pregnant      glucose      pressure      triceps      insulin      mass      pedigree
##      4.865672  141.257463   70.824627   22.164179  100.335821   35.142537   0.550500
##           age
##      37.067164
```



```
colMeans(group2)
```

```
##    pregnant    glucose    pressure    triceps    insulin    mass    pedigree
##    3.298000  109.980000   68.184000   19.664000   68.792000   30.304200   0.429734
##          age
##    31.190000
```

```
T1 = 0.5*log(det(cov(group2))/det(cov(group1)))
```

```
T2 = 0.5*(t(colMeans(group2))%*%solve(cov(group2))%*%(colMeans(group2))-t(colMeans(group1))%*%solve(cov
```

```
db = T1+T2
```

```
colMeans(group1)%*%solve(cov(group1))
```

```
##    pregnant    glucose    pressure    triceps    insulin    mass    pedigree
## [1,] 0.1729652 0.1225318 0.06969635 -0.02998349 -0.004515142 0.7182942 3.113763
##          age
## [1,] 0.3064803
```

```
colMeans(group2)%*%solve(cov(group2))
```

```
##    pregnant    glucose    pressure    triceps    insulin    mass    pedigree
## [1,] -0.1395849 0.1428478 0.09828429 0.02951318 -0.01812551 0.3833575 3.676697
##          age
## [1,] 0.1143517
```

```
solve(cov(group2))-solve(cov(group1))
```

```
##          pregnant    glucose    pressure    triceps    insulin
## pregnant  7.509728e-02 -0.0003835994  2.501331e-05  9.505874e-04 -1.178500e-04
## glucose  -3.835994e-04  0.0008279548 -2.861588e-04  2.929996e-04 -1.459817e-04
## pressure  2.501331e-05 -0.0002861588  1.252018e-03  2.373027e-04  3.064438e-05
## triceps   9.505874e-04  0.0002929996  2.373027e-04  1.666327e-03 -7.591392e-05
## insulin  -1.178500e-04 -0.0001459817  3.064438e-05 -7.591392e-05  8.476254e-05
## mass     -4.335091e-03  0.0001103101 -1.837359e-03 -1.586684e-03 -2.823432e-04
## pedigree  1.491800e-01  0.0036112715 -5.853952e-04  4.328708e-02 -9.069247e-03
## age      -1.122608e-02 -0.0006914890  4.403570e-04  1.459966e-04  2.829777e-04
##          mass    pedigree    age
## pregnant -0.0043350913  0.1491800230 -0.0112260840
## glucose   0.0001103101  0.0036112715 -0.0006914890
## pressure -0.0018373588 -0.0005853952  0.0004403570
## triceps  -0.0015866837  0.0432870830  0.0001459966
## insulin  -0.0002823432 -0.0090692473  0.0002829777
## mass      0.0011492313  0.0129252447 -0.0028147904
## pedigree  0.0129252447  4.1673558045 -0.0609225598
## age       -0.0028147904 -0.0609225598  0.0007643262
```

```
print(db)
```

```
##          [,1]
## [1,] -12.5319
```

QDA (split) (cont.) - Confusion Matrix

```
library(mlbench)
```

```
library(MASS)
```

```
library(caret)
```

```
data(PimaIndiansDiabetes)
```

```

# convert to a factor
PimaIndiansDiabetes$diabetes <- as.factor(PimaIndiansDiabetes$diabetes)

# numeric data
PimaIndiansDiabetes <- PimaIndiansDiabetes[, sapply(PimaIndiansDiabetes, is.numeric) | colnames(PimaIndiansDiabetes) == "diabetes"]

# training and test sets
set.seed(123)
train_Index <- createDataPartition(PimaIndiansDiabetes$diabetes, p = .7,
                                   list = FALSE,
                                   times = 1)
Pima_Train <- PimaIndiansDiabetes[ train_Index,]
PimaTest  <- PimaIndiansDiabetes[-train_Index,]

# train QDA model
qda_model <- qda(diabetes ~ ., data = Pima_Train)

# predictions on test set
qda_predictions <- predict(qda_model, PimaTest)

# evaluate model
conf_matrix <- confusionMatrix(qda_predictions$class, PimaTest$diabetes)
print(conf_matrix)

```

```

## Confusion Matrix and Statistics
##
##           Reference
## Prediction neg pos
##           neg 122  39
##           pos  28  41
##
##               Accuracy : 0.7087
##               95% CI : (0.6454, 0.7666)
##           No Information Rate : 0.6522
##           P-Value [Acc > NIR] : 0.04036
##
##               Kappa : 0.3366
##
##  Mcnemar's Test P-Value : 0.22182
##
##           Sensitivity : 0.8133
##           Specificity : 0.5125
##           Pos Pred Value : 0.7578
##           Neg Pred Value : 0.5942
##           Prevalence : 0.6522
##           Detection Rate : 0.5304
##           Detection Prevalence : 0.7000
##           Balanced Accuracy : 0.6629
##
##           'Positive' Class : neg
##

```

From the confusion matrix, we see that it yielded 122 true negatives and 28 false negatives, and 41 true positives and 39 false positives. False positives are indicative of Type I error, and false negatives are indicative

of Type II error. This means from our confusion matrix, 39 healthy women were diagnosed with onset diabetes, and 28 women with onset diabetes were incorrectly diagnosed as healthy.

QDA (split) Performance Metrics

```
# performance metrics
accuracy <- conf_matrix$overall['Accuracy']
precision <- conf_matrix$byClass['Pos Pred Value']
recall <- conf_matrix$byClass['Sensitivity']
f1 <- 2 * (precision * recall) / (precision + recall)

cat("Accuracy of QDA Model: ", accuracy, "\n")
```

```
## Accuracy of QDA Model: 0.7086957
```

```
cat("Precision of QDA Model: ", precision, "\n")
```

```
## Precision of QDA Model: 0.757764
```

```
cat("Recall of QDA Model: ", recall, "\n")
```

```
## Recall of QDA Model: 0.8133333
```

```
cat("F1 Score of QDA Model: ", f1, "\n")
```

```
## F1 Score of QDA Model: 0.7845659
```

```
# misclassification error rate
total <- sum(conf_matrix$table)
misclassifications <- total - sum(diag(conf_matrix$table))
error_rate <- misclassifications / total

cat("Misclassification Error Rate: ", error_rate, "\n")
```

```
## Misclassification Error Rate: 0.2913043
```

From this table, we can identify several key findings. First, we obtain an accuracy rate of 70.87%, which is the percentage of correctly classified instances. We get a precision rate of 75.78%, which is the percentage of positive identifications that actually prove to be correct. The recall percentage is 81.33%, the percentage of actual positives that are correctly identified. The F1 score, which is the mean of precision and recall, is 78.46%. Finally, we obtain a misclassification error rate of 29.13%.

For precision, recall, and F1 score, we ideally would desire rates at 80% or above. Our precision and F1 rates are somewhat close to 80%, while our recall percentage is over 80%, which is generally good. We want a high accuracy, ideally above 90%, but our accuracy is at 70.87%, which is not good enough. Our misclassification error rate is mediocre at 29.13%, with a rate at 40% or above being considered poor.

QDA with cross-validation - Confusion Matrix

```
data(PimaIndiansDiabetes)
PimaIndiansDiabetes$diabetes <- as.factor(PimaIndiansDiabetes$diabetes)
PimaIndiansDiabetes <- PimaIndiansDiabetes[, sapply(PimaIndiansDiabetes, is.numeric) | colnames(PimaIndiansDiabetes) != "diabetes"]

# train QDA model
control <- trainControl(method = "cv", number = 10)
set.seed(123)
qda_model2 <- train(diabetes ~ ., data = PimaIndiansDiabetes, method = "qda", trControl = control)
print(qda_model2)
```

```
## Quadratic Discriminant Analysis
##
```

```

## 768 samples
## 8 predictor
## 2 classes: 'neg', 'pos'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 691, 691, 691, 692, 692, 691, ...
## Resampling results:
##
## Accuracy Kappa
## 0.7409091 0.4097772

# predictions
qda_predictions2 <- predict(qda_model2, PimaIndiansDiabetes)
qda_probs2 <- predict(qda_model2, PimaIndiansDiabetes, type = "prob")

# confusion matrix
conf_matrix2 <- confusionMatrix(qda_predictions2, PimaIndiansDiabetes$diabetes)
print(conf_matrix2)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction neg pos
##      neg 432 113
##      pos  68 155
##
##              Accuracy : 0.7643
##              95% CI : (0.7327, 0.7939)
##      No Information Rate : 0.651
##      P-Value [Acc > NIR] : 7.13e-12
##
##              Kappa : 0.4603
##
##  Mcnemar's Test P-Value : 0.001074
##
##              Sensitivity : 0.8640
##              Specificity : 0.5784
##      Pos Pred Value : 0.7927
##      Neg Pred Value : 0.6951
##              Prevalence : 0.6510
##      Detection Rate : 0.5625
##      Detection Prevalence : 0.7096
##      Balanced Accuracy : 0.7212
##
##      'Positive' Class : neg
##

```

Examining the confusion matrix from the QDA model with cross-validation, we have 432 true negatives and 68 false negatives, and 155 true positives and 113 false positives. This indicates that 113 healthy women were incorrectly diagnosed with diabetes (Type I error), and 68 women with diabetes were incorrectly diagnosed as healthy (Type II error).

QDA with cross-validation - Performance Metrics

```

# misclassification error rate
error_rate <- 1 - conf_matrix2$overall['Accuracy']
cat("Misclassification Error Rate: ", error_rate, "\n")

## Misclassification Error Rate: 0.2356771

# performance metrics
accuracy <- conf_matrix2$overall['Accuracy']
precision <- conf_matrix2$byClass['Pos Pred Value']
recall <- conf_matrix2$byClass['Sensitivity']
f1_score <- conf_matrix2$byClass['F1']

cat("Accuracy of QDA Model w/ Cross-Validation: ", accuracy, "\n")

## Accuracy of QDA Model w/ Cross-Validation: 0.7643229
cat("Precision of QDA Model w/ Cross-Validation: ", precision, "\n")

## Precision of QDA Model w/ Cross-Validation: 0.7926606
cat("Recall of QDA Model w/ Cross-Validation: ", recall, "\n")

## Recall of QDA Model w/ Cross-Validation: 0.864
cat("F1 Score of QDA Model w/ Cross-Validation: ", f1_score, "\n")

## F1 Score of QDA Model w/ Cross-Validation: 0.8267943

```

From running a QDA model with cross-validation, we obtain several interesting findings. First, we have an accuracy of 76.43%, which is an improvement over the prior model WITH training and test sets. We also obtain improved precision and recall rates at 79.27% and 86.4%, respectively, an improved F1 score of 82.68%, and a lower misclassification error rate of 23.57%.

Our accuracy rate is about 6% closer to a generally “good” rate of 80%, but still not at our desired “great” rate of 90% or higher. Our precision rate is almost at 80%, which we can consider generally good, and our recall rate is sitting at a very good percentage. Our F1 score is also now above 80%, which we can also now consider to be good. The misclassification error rate did not drop as much as we wanted it to, with it still sitting at a somewhat mediocre percentage, but it is still beneficial that it lowered at all.

QDA without splitting

```

library(mlbench)
library(MASS)
library(caret)
data(PimaIndiansDiabetes)

# convert diabetes column to a factor
PimaIndiansDiabetes$diabetes <- as.factor(PimaIndiansDiabetes$diabetes)

# numeric data
PimaIndiansDiabetes <- PimaIndiansDiabetes[, sapply(PimaIndiansDiabetes, is.numeric) | colnames(PimaIndiansDiabetes) != "diabetes"]

# train QDA model
qda_no_split <- qda(diabetes ~ ., data = PimaIndiansDiabetes)

# predictions on entire set
qda_predictions3 <- predict(qda_no_split, PimaIndiansDiabetes)

# evaluate model

```

```
conf_matrix3 <- confusionMatrix(qda_predictions3$class, PimaIndiansDiabetes$diabetes)
print(conf_matrix3)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction neg pos
##           neg 432 113
##           pos  68 155
##
##           Accuracy : 0.7643
##           95% CI : (0.7327, 0.7939)
##           No Information Rate : 0.651
##           P-Value [Acc > NIR] : 7.13e-12
##
##           Kappa : 0.4603
##
## Mcnemar's Test P-Value : 0.001074
##
##           Sensitivity : 0.8640
##           Specificity : 0.5784
##           Pos Pred Value : 0.7927
##           Neg Pred Value : 0.6951
##           Prevalence : 0.6510
##           Detection Rate : 0.5625
##           Detection Prevalence : 0.7096
##           Balanced Accuracy : 0.7212
##
##           'Positive' Class : neg
##
```

QDA without splitting - Performance Metrics

```
# misclassification error rate
error_rate <- 1 - conf_matrix3$overall['Accuracy']
cat("Misclassification Error Rate: ", error_rate, "\n")
```

```
## Misclassification Error Rate: 0.2356771
```

```
# performance metrics
accuracy <- conf_matrix3$overall['Accuracy']
precision <- conf_matrix3$byClass['Pos Pred Value']
recall <- conf_matrix3$byClass['Sensitivity']
f1_score <- conf_matrix3$byClass['F1']

cat("Accuracy of QDA Model w/ Cross-Validation: ", accuracy, "\n")
```

```
## Accuracy of QDA Model w/ Cross-Validation: 0.7643229
```

```
cat("Precision of QDA Model w/ Cross-Validation: ", precision, "\n")
```

```
## Precision of QDA Model w/ Cross-Validation: 0.7926606
```

```
cat("Recall of QDA Model w/ Cross-Validation: ", recall, "\n")
```

```
## Recall of QDA Model w/ Cross-Validation: 0.864
```

```
cat("F1 Score of QDA Model w/ Cross-Validation: ", f1_score, "\n")
```

```
## F1 Score of QDA Model w/ Cross-Validation: 0.8267943
```

Since both the QDA cross-validation model and the QDA model without splitting result in the same exact accuracy, precision, recall, F1 score, and misclassification error rate, the cross validation model is better because it has less risk of overfitting, so it will make more accurate predictions on new data compared to the model without splitting data.

Analysis of QDA Results:

One of the main findings from our work with QDA (quantitative descriptive analysis) was that every metric that deals with accuracy and performance improved after doing QDA without splitting and k-fold cross-validation on the data. In other words, the metrics improved after switching to either QDA with k-fold cross-validation or QDA without splitting. Now, we would generally conclude that the efficiency in QDA lies in simply leaving out the process of splitting the data, but there is also the concern of overfitting the data and evaluation bias, two things that almost necessitate splitting the data for QDA. Overfitting the data occurs when the model learns the training data so well that it results in higher accuracy, but with poor generalization on newer data. Evaluation bias occurs when evaluation is performed on the same data it was trained on, and since the model has already been exposed to the data during the process of training, the performance does not accurately reflect how it would on newer data.

Splitting the data allows us to evaluate how the model performs on separate and unseen subsets of data, which is far more useful than subsets of data the model has already gotten accustomed to. Further, the key for evaluating a model also lies in how it works on newer data, because we want to see how it may perform in other real-world scenarios. With these factors, we cannot conclude that performing QDA without splitting the data into test and training sets. The best method would be QDA with k-fold cross-validation, because it splits into ten folds and evaluates for each fold, making the accuracy of this model more legitimate and better fitting for future subsets of data.

PCA

```
#PCA
```

```
library(ade4)
```

```
data(PimaIndiansDiabetes)
PimaIndiansDiabetes
```

##	pregnant	glucose	pressure	triceps	insulin	mass	pedigree	age	diabetes
## 1	6	148	72	35	0	33.6	0.627	50	pos
## 2	1	85	66	29	0	26.6	0.351	31	neg
## 3	8	183	64	0	0	23.3	0.672	32	pos
## 4	1	89	66	23	94	28.1	0.167	21	neg
## 5	0	137	40	35	168	43.1	2.288	33	pos
## 6	5	116	74	0	0	25.6	0.201	30	neg
## 7	3	78	50	32	88	31.0	0.248	26	pos
## 8	10	115	0	0	0	35.3	0.134	29	neg
## 9	2	197	70	45	543	30.5	0.158	53	pos
## 10	8	125	96	0	0	0.0	0.232	54	pos
## 11	4	110	92	0	0	37.6	0.191	30	neg
## 12	10	168	74	0	0	38.0	0.537	34	pos
## 13	10	139	80	0	0	27.1	1.441	57	neg
## 14	1	189	60	23	846	30.1	0.398	59	pos
## 15	5	166	72	19	175	25.8	0.587	51	pos
## 16	7	100	0	0	0	30.0	0.484	32	pos
## 17	0	118	84	47	230	45.8	0.551	31	pos

## 18	7	107	74	0	0	29.6	0.254	31	pos
## 19	1	103	30	38	83	43.3	0.183	33	neg
## 20	1	115	70	30	96	34.6	0.529	32	pos
## 21	3	126	88	41	235	39.3	0.704	27	neg
## 22	8	99	84	0	0	35.4	0.388	50	neg
## 23	7	196	90	0	0	39.8	0.451	41	pos
## 24	9	119	80	35	0	29.0	0.263	29	pos
## 25	11	143	94	33	146	36.6	0.254	51	pos
## 26	10	125	70	26	115	31.1	0.205	41	pos
## 27	7	147	76	0	0	39.4	0.257	43	pos
## 28	1	97	66	15	140	23.2	0.487	22	neg
## 29	13	145	82	19	110	22.2	0.245	57	neg
## 30	5	117	92	0	0	34.1	0.337	38	neg
## 31	5	109	75	26	0	36.0	0.546	60	neg
## 32	3	158	76	36	245	31.6	0.851	28	pos
## 33	3	88	58	11	54	24.8	0.267	22	neg
## 34	6	92	92	0	0	19.9	0.188	28	neg
## 35	10	122	78	31	0	27.6	0.512	45	neg
## 36	4	103	60	33	192	24.0	0.966	33	neg
## 37	11	138	76	0	0	33.2	0.420	35	neg
## 38	9	102	76	37	0	32.9	0.665	46	pos
## 39	2	90	68	42	0	38.2	0.503	27	pos
## 40	4	111	72	47	207	37.1	1.390	56	pos
## 41	3	180	64	25	70	34.0	0.271	26	neg
## 42	7	133	84	0	0	40.2	0.696	37	neg
## 43	7	106	92	18	0	22.7	0.235	48	neg
## 44	9	171	110	24	240	45.4	0.721	54	pos
## 45	7	159	64	0	0	27.4	0.294	40	neg
## 46	0	180	66	39	0	42.0	1.893	25	pos
## 47	1	146	56	0	0	29.7	0.564	29	neg
## 48	2	71	70	27	0	28.0	0.586	22	neg
## 49	7	103	66	32	0	39.1	0.344	31	pos
## 50	7	105	0	0	0	0.0	0.305	24	neg
## 51	1	103	80	11	82	19.4	0.491	22	neg
## 52	1	101	50	15	36	24.2	0.526	26	neg
## 53	5	88	66	21	23	24.4	0.342	30	neg
## 54	8	176	90	34	300	33.7	0.467	58	pos
## 55	7	150	66	42	342	34.7	0.718	42	neg
## 56	1	73	50	10	0	23.0	0.248	21	neg
## 57	7	187	68	39	304	37.7	0.254	41	pos
## 58	0	100	88	60	110	46.8	0.962	31	neg
## 59	0	146	82	0	0	40.5	1.781	44	neg
## 60	0	105	64	41	142	41.5	0.173	22	neg
## 61	2	84	0	0	0	0.0	0.304	21	neg
## 62	8	133	72	0	0	32.9	0.270	39	pos
## 63	5	44	62	0	0	25.0	0.587	36	neg
## 64	2	141	58	34	128	25.4	0.699	24	neg
## 65	7	114	66	0	0	32.8	0.258	42	pos
## 66	5	99	74	27	0	29.0	0.203	32	neg
## 67	0	109	88	30	0	32.5	0.855	38	pos
## 68	2	109	92	0	0	42.7	0.845	54	neg
## 69	1	95	66	13	38	19.6	0.334	25	neg
## 70	4	146	85	27	100	28.9	0.189	27	neg
## 71	2	100	66	20	90	32.9	0.867	28	pos

## 72	5	139	64	35	140	28.6	0.411	26	neg
## 73	13	126	90	0	0	43.4	0.583	42	pos
## 74	4	129	86	20	270	35.1	0.231	23	neg
## 75	1	79	75	30	0	32.0	0.396	22	neg
## 76	1	0	48	20	0	24.7	0.140	22	neg
## 77	7	62	78	0	0	32.6	0.391	41	neg
## 78	5	95	72	33	0	37.7	0.370	27	neg
## 79	0	131	0	0	0	43.2	0.270	26	pos
## 80	2	112	66	22	0	25.0	0.307	24	neg
## 81	3	113	44	13	0	22.4	0.140	22	neg
## 82	2	74	0	0	0	0.0	0.102	22	neg
## 83	7	83	78	26	71	29.3	0.767	36	neg
## 84	0	101	65	28	0	24.6	0.237	22	neg
## 85	5	137	108	0	0	48.8	0.227	37	pos
## 86	2	110	74	29	125	32.4	0.698	27	neg
## 87	13	106	72	54	0	36.6	0.178	45	neg
## 88	2	100	68	25	71	38.5	0.324	26	neg
## 89	15	136	70	32	110	37.1	0.153	43	pos
## 90	1	107	68	19	0	26.5	0.165	24	neg
## 91	1	80	55	0	0	19.1	0.258	21	neg
## 92	4	123	80	15	176	32.0	0.443	34	neg
## 93	7	81	78	40	48	46.7	0.261	42	neg
## 94	4	134	72	0	0	23.8	0.277	60	pos
## 95	2	142	82	18	64	24.7	0.761	21	neg
## 96	6	144	72	27	228	33.9	0.255	40	neg
## 97	2	92	62	28	0	31.6	0.130	24	neg
## 98	1	71	48	18	76	20.4	0.323	22	neg
## 99	6	93	50	30	64	28.7	0.356	23	neg
## 100	1	122	90	51	220	49.7	0.325	31	pos
## 101	1	163	72	0	0	39.0	1.222	33	pos
## 102	1	151	60	0	0	26.1	0.179	22	neg
## 103	0	125	96	0	0	22.5	0.262	21	neg
## 104	1	81	72	18	40	26.6	0.283	24	neg
## 105	2	85	65	0	0	39.6	0.930	27	neg
## 106	1	126	56	29	152	28.7	0.801	21	neg
## 107	1	96	122	0	0	22.4	0.207	27	neg
## 108	4	144	58	28	140	29.5	0.287	37	neg
## 109	3	83	58	31	18	34.3	0.336	25	neg
## 110	0	95	85	25	36	37.4	0.247	24	pos
## 111	3	171	72	33	135	33.3	0.199	24	pos
## 112	8	155	62	26	495	34.0	0.543	46	pos
## 113	1	89	76	34	37	31.2	0.192	23	neg
## 114	4	76	62	0	0	34.0	0.391	25	neg
## 115	7	160	54	32	175	30.5	0.588	39	pos
## 116	4	146	92	0	0	31.2	0.539	61	pos
## 117	5	124	74	0	0	34.0	0.220	38	pos
## 118	5	78	48	0	0	33.7	0.654	25	neg
## 119	4	97	60	23	0	28.2	0.443	22	neg
## 120	4	99	76	15	51	23.2	0.223	21	neg
## 121	0	162	76	56	100	53.2	0.759	25	pos
## 122	6	111	64	39	0	34.2	0.260	24	neg
## 123	2	107	74	30	100	33.6	0.404	23	neg
## 124	5	132	80	0	0	26.8	0.186	69	neg
## 125	0	113	76	0	0	33.3	0.278	23	pos

## 126	1	88	30	42	99 55.0	0.496	26	pos
## 127	3	120	70	30	135 42.9	0.452	30	neg
## 128	1	118	58	36	94 33.3	0.261	23	neg
## 129	1	117	88	24	145 34.5	0.403	40	pos
## 130	0	105	84	0	0 27.9	0.741	62	pos
## 131	4	173	70	14	168 29.7	0.361	33	pos
## 132	9	122	56	0	0 33.3	1.114	33	pos
## 133	3	170	64	37	225 34.5	0.356	30	pos
## 134	8	84	74	31	0 38.3	0.457	39	neg
## 135	2	96	68	13	49 21.1	0.647	26	neg
## 136	2	125	60	20	140 33.8	0.088	31	neg
## 137	0	100	70	26	50 30.8	0.597	21	neg
## 138	0	93	60	25	92 28.7	0.532	22	neg
## 139	0	129	80	0	0 31.2	0.703	29	neg
## 140	5	105	72	29	325 36.9	0.159	28	neg
## 141	3	128	78	0	0 21.1	0.268	55	neg
## 142	5	106	82	30	0 39.5	0.286	38	neg
## 143	2	108	52	26	63 32.5	0.318	22	neg
## 144	10	108	66	0	0 32.4	0.272	42	pos
## 145	4	154	62	31	284 32.8	0.237	23	neg
## 146	0	102	75	23	0 0.0	0.572	21	neg
## 147	9	57	80	37	0 32.8	0.096	41	neg
## 148	2	106	64	35	119 30.5	1.400	34	neg
## 149	5	147	78	0	0 33.7	0.218	65	neg
## 150	2	90	70	17	0 27.3	0.085	22	neg
## 151	1	136	74	50	204 37.4	0.399	24	neg
## 152	4	114	65	0	0 21.9	0.432	37	neg
## 153	9	156	86	28	155 34.3	1.189	42	pos
## 154	1	153	82	42	485 40.6	0.687	23	neg
## 155	8	188	78	0	0 47.9	0.137	43	pos
## 156	7	152	88	44	0 50.0	0.337	36	pos
## 157	2	99	52	15	94 24.6	0.637	21	neg
## 158	1	109	56	21	135 25.2	0.833	23	neg
## 159	2	88	74	19	53 29.0	0.229	22	neg
## 160	17	163	72	41	114 40.9	0.817	47	pos
## 161	4	151	90	38	0 29.7	0.294	36	neg
## 162	7	102	74	40	105 37.2	0.204	45	neg
## 163	0	114	80	34	285 44.2	0.167	27	neg
## 164	2	100	64	23	0 29.7	0.368	21	neg
## 165	0	131	88	0	0 31.6	0.743	32	pos
## 166	6	104	74	18	156 29.9	0.722	41	pos
## 167	3	148	66	25	0 32.5	0.256	22	neg
## 168	4	120	68	0	0 29.6	0.709	34	neg
## 169	4	110	66	0	0 31.9	0.471	29	neg
## 170	3	111	90	12	78 28.4	0.495	29	neg
## 171	6	102	82	0	0 30.8	0.180	36	pos
## 172	6	134	70	23	130 35.4	0.542	29	pos
## 173	2	87	0	23	0 28.9	0.773	25	neg
## 174	1	79	60	42	48 43.5	0.678	23	neg
## 175	2	75	64	24	55 29.7	0.370	33	neg
## 176	8	179	72	42	130 32.7	0.719	36	pos
## 177	6	85	78	0	0 31.2	0.382	42	neg
## 178	0	129	110	46	130 67.1	0.319	26	pos
## 179	5	143	78	0	0 45.0	0.190	47	neg

## 180	5	130	82	0	0 39.1	0.956	37	pos
## 181	6	87	80	0	0 23.2	0.084	32	neg
## 182	0	119	64	18	92 34.9	0.725	23	neg
## 183	1	0	74	20	23 27.7	0.299	21	neg
## 184	5	73	60	0	0 26.8	0.268	27	neg
## 185	4	141	74	0	0 27.6	0.244	40	neg
## 186	7	194	68	28	0 35.9	0.745	41	pos
## 187	8	181	68	36	495 30.1	0.615	60	pos
## 188	1	128	98	41	58 32.0	1.321	33	pos
## 189	8	109	76	39	114 27.9	0.640	31	pos
## 190	5	139	80	35	160 31.6	0.361	25	pos
## 191	3	111	62	0	0 22.6	0.142	21	neg
## 192	9	123	70	44	94 33.1	0.374	40	neg
## 193	7	159	66	0	0 30.4	0.383	36	pos
## 194	11	135	0	0	0 52.3	0.578	40	pos
## 195	8	85	55	20	0 24.4	0.136	42	neg
## 196	5	158	84	41	210 39.4	0.395	29	pos
## 197	1	105	58	0	0 24.3	0.187	21	neg
## 198	3	107	62	13	48 22.9	0.678	23	pos
## 199	4	109	64	44	99 34.8	0.905	26	pos
## 200	4	148	60	27	318 30.9	0.150	29	pos
## 201	0	113	80	16	0 31.0	0.874	21	neg
## 202	1	138	82	0	0 40.1	0.236	28	neg
## 203	0	108	68	20	0 27.3	0.787	32	neg
## 204	2	99	70	16	44 20.4	0.235	27	neg
## 205	6	103	72	32	190 37.7	0.324	55	neg
## 206	5	111	72	28	0 23.9	0.407	27	neg
## 207	8	196	76	29	280 37.5	0.605	57	pos
## 208	5	162	104	0	0 37.7	0.151	52	pos
## 209	1	96	64	27	87 33.2	0.289	21	neg
## 210	7	184	84	33	0 35.5	0.355	41	pos
## 211	2	81	60	22	0 27.7	0.290	25	neg
## 212	0	147	85	54	0 42.8	0.375	24	neg
## 213	7	179	95	31	0 34.2	0.164	60	neg
## 214	0	140	65	26	130 42.6	0.431	24	pos
## 215	9	112	82	32	175 34.2	0.260	36	pos
## 216	12	151	70	40	271 41.8	0.742	38	pos
## 217	5	109	62	41	129 35.8	0.514	25	pos
## 218	6	125	68	30	120 30.0	0.464	32	neg
## 219	5	85	74	22	0 29.0	1.224	32	pos
## 220	5	112	66	0	0 37.8	0.261	41	pos
## 221	0	177	60	29	478 34.6	1.072	21	pos
## 222	2	158	90	0	0 31.6	0.805	66	pos
## 223	7	119	0	0	0 25.2	0.209	37	neg
## 224	7	142	60	33	190 28.8	0.687	61	neg
## 225	1	100	66	15	56 23.6	0.666	26	neg
## 226	1	87	78	27	32 34.6	0.101	22	neg
## 227	0	101	76	0	0 35.7	0.198	26	neg
## 228	3	162	52	38	0 37.2	0.652	24	pos
## 229	4	197	70	39	744 36.7	2.329	31	neg
## 230	0	117	80	31	53 45.2	0.089	24	neg
## 231	4	142	86	0	0 44.0	0.645	22	pos
## 232	6	134	80	37	370 46.2	0.238	46	pos
## 233	1	79	80	25	37 25.4	0.583	22	neg

## 234	4	122	68	0	0 35.0	0.394	29	neg
## 235	3	74	68	28	45 29.7	0.293	23	neg
## 236	4	171	72	0	0 43.6	0.479	26	pos
## 237	7	181	84	21	192 35.9	0.586	51	pos
## 238	0	179	90	27	0 44.1	0.686	23	pos
## 239	9	164	84	21	0 30.8	0.831	32	pos
## 240	0	104	76	0	0 18.4	0.582	27	neg
## 241	1	91	64	24	0 29.2	0.192	21	neg
## 242	4	91	70	32	88 33.1	0.446	22	neg
## 243	3	139	54	0	0 25.6	0.402	22	pos
## 244	6	119	50	22	176 27.1	1.318	33	pos
## 245	2	146	76	35	194 38.2	0.329	29	neg
## 246	9	184	85	15	0 30.0	1.213	49	pos
## 247	10	122	68	0	0 31.2	0.258	41	neg
## 248	0	165	90	33	680 52.3	0.427	23	neg
## 249	9	124	70	33	402 35.4	0.282	34	neg
## 250	1	111	86	19	0 30.1	0.143	23	neg
## 251	9	106	52	0	0 31.2	0.380	42	neg
## 252	2	129	84	0	0 28.0	0.284	27	neg
## 253	2	90	80	14	55 24.4	0.249	24	neg
## 254	0	86	68	32	0 35.8	0.238	25	neg
## 255	12	92	62	7	258 27.6	0.926	44	pos
## 256	1	113	64	35	0 33.6	0.543	21	pos
## 257	3	111	56	39	0 30.1	0.557	30	neg
## 258	2	114	68	22	0 28.7	0.092	25	neg
## 259	1	193	50	16	375 25.9	0.655	24	neg
## 260	11	155	76	28	150 33.3	1.353	51	pos
## 261	3	191	68	15	130 30.9	0.299	34	neg
## 262	3	141	0	0	0 30.0	0.761	27	pos
## 263	4	95	70	32	0 32.1	0.612	24	neg
## 264	3	142	80	15	0 32.4	0.200	63	neg
## 265	4	123	62	0	0 32.0	0.226	35	pos
## 266	5	96	74	18	67 33.6	0.997	43	neg
## 267	0	138	0	0	0 36.3	0.933	25	pos
## 268	2	128	64	42	0 40.0	1.101	24	neg
## 269	0	102	52	0	0 25.1	0.078	21	neg
## 270	2	146	0	0	0 27.5	0.240	28	pos
## 271	10	101	86	37	0 45.6	1.136	38	pos
## 272	2	108	62	32	56 25.2	0.128	21	neg
## 273	3	122	78	0	0 23.0	0.254	40	neg
## 274	1	71	78	50	45 33.2	0.422	21	neg
## 275	13	106	70	0	0 34.2	0.251	52	neg
## 276	2	100	70	52	57 40.5	0.677	25	neg
## 277	7	106	60	24	0 26.5	0.296	29	pos
## 278	0	104	64	23	116 27.8	0.454	23	neg
## 279	5	114	74	0	0 24.9	0.744	57	neg
## 280	2	108	62	10	278 25.3	0.881	22	neg
## 281	0	146	70	0	0 37.9	0.334	28	pos
## 282	10	129	76	28	122 35.9	0.280	39	neg
## 283	7	133	88	15	155 32.4	0.262	37	neg
## 284	7	161	86	0	0 30.4	0.165	47	pos
## 285	2	108	80	0	0 27.0	0.259	52	pos
## 286	7	136	74	26	135 26.0	0.647	51	neg
## 287	5	155	84	44	545 38.7	0.619	34	neg

## 288	1	119	86	39	220	45.6	0.808	29	pos
## 289	4	96	56	17	49	20.8	0.340	26	neg
## 290	5	108	72	43	75	36.1	0.263	33	neg
## 291	0	78	88	29	40	36.9	0.434	21	neg
## 292	0	107	62	30	74	36.6	0.757	25	pos
## 293	2	128	78	37	182	43.3	1.224	31	pos
## 294	1	128	48	45	194	40.5	0.613	24	pos
## 295	0	161	50	0	0	21.9	0.254	65	neg
## 296	6	151	62	31	120	35.5	0.692	28	neg
## 297	2	146	70	38	360	28.0	0.337	29	pos
## 298	0	126	84	29	215	30.7	0.520	24	neg
## 299	14	100	78	25	184	36.6	0.412	46	pos
## 300	8	112	72	0	0	23.6	0.840	58	neg
## 301	0	167	0	0	0	32.3	0.839	30	pos
## 302	2	144	58	33	135	31.6	0.422	25	pos
## 303	5	77	82	41	42	35.8	0.156	35	neg
## 304	5	115	98	0	0	52.9	0.209	28	pos
## 305	3	150	76	0	0	21.0	0.207	37	neg
## 306	2	120	76	37	105	39.7	0.215	29	neg
## 307	10	161	68	23	132	25.5	0.326	47	pos
## 308	0	137	68	14	148	24.8	0.143	21	neg
## 309	0	128	68	19	180	30.5	1.391	25	pos
## 310	2	124	68	28	205	32.9	0.875	30	pos
## 311	6	80	66	30	0	26.2	0.313	41	neg
## 312	0	106	70	37	148	39.4	0.605	22	neg
## 313	2	155	74	17	96	26.6	0.433	27	pos
## 314	3	113	50	10	85	29.5	0.626	25	neg
## 315	7	109	80	31	0	35.9	1.127	43	pos
## 316	2	112	68	22	94	34.1	0.315	26	neg
## 317	3	99	80	11	64	19.3	0.284	30	neg
## 318	3	182	74	0	0	30.5	0.345	29	pos
## 319	3	115	66	39	140	38.1	0.150	28	neg
## 320	6	194	78	0	0	23.5	0.129	59	pos
## 321	4	129	60	12	231	27.5	0.527	31	neg
## 322	3	112	74	30	0	31.6	0.197	25	pos
## 323	0	124	70	20	0	27.4	0.254	36	pos
## 324	13	152	90	33	29	26.8	0.731	43	pos
## 325	2	112	75	32	0	35.7	0.148	21	neg
## 326	1	157	72	21	168	25.6	0.123	24	neg
## 327	1	122	64	32	156	35.1	0.692	30	pos
## 328	10	179	70	0	0	35.1	0.200	37	neg
## 329	2	102	86	36	120	45.5	0.127	23	pos
## 330	6	105	70	32	68	30.8	0.122	37	neg
## 331	8	118	72	19	0	23.1	1.476	46	neg
## 332	2	87	58	16	52	32.7	0.166	25	neg
## 333	1	180	0	0	0	43.3	0.282	41	pos
## 334	12	106	80	0	0	23.6	0.137	44	neg
## 335	1	95	60	18	58	23.9	0.260	22	neg
## 336	0	165	76	43	255	47.9	0.259	26	neg
## 337	0	117	0	0	0	33.8	0.932	44	neg
## 338	5	115	76	0	0	31.2	0.343	44	pos
## 339	9	152	78	34	171	34.2	0.893	33	pos
## 340	7	178	84	0	0	39.9	0.331	41	pos
## 341	1	130	70	13	105	25.9	0.472	22	neg

## 342	1	95	74	21	73	25.9	0.673	36	neg
## 343	1	0	68	35	0	32.0	0.389	22	neg
## 344	5	122	86	0	0	34.7	0.290	33	neg
## 345	8	95	72	0	0	36.8	0.485	57	neg
## 346	8	126	88	36	108	38.5	0.349	49	neg
## 347	1	139	46	19	83	28.7	0.654	22	neg
## 348	3	116	0	0	0	23.5	0.187	23	neg
## 349	3	99	62	19	74	21.8	0.279	26	neg
## 350	5	0	80	32	0	41.0	0.346	37	pos
## 351	4	92	80	0	0	42.2	0.237	29	neg
## 352	4	137	84	0	0	31.2	0.252	30	neg
## 353	3	61	82	28	0	34.4	0.243	46	neg
## 354	1	90	62	12	43	27.2	0.580	24	neg
## 355	3	90	78	0	0	42.7	0.559	21	neg
## 356	9	165	88	0	0	30.4	0.302	49	pos
## 357	1	125	50	40	167	33.3	0.962	28	pos
## 358	13	129	0	30	0	39.9	0.569	44	pos
## 359	12	88	74	40	54	35.3	0.378	48	neg
## 360	1	196	76	36	249	36.5	0.875	29	pos
## 361	5	189	64	33	325	31.2	0.583	29	pos
## 362	5	158	70	0	0	29.8	0.207	63	neg
## 363	5	103	108	37	0	39.2	0.305	65	neg
## 364	4	146	78	0	0	38.5	0.520	67	pos
## 365	4	147	74	25	293	34.9	0.385	30	neg
## 366	5	99	54	28	83	34.0	0.499	30	neg
## 367	6	124	72	0	0	27.6	0.368	29	pos
## 368	0	101	64	17	0	21.0	0.252	21	neg
## 369	3	81	86	16	66	27.5	0.306	22	neg
## 370	1	133	102	28	140	32.8	0.234	45	pos
## 371	3	173	82	48	465	38.4	2.137	25	pos
## 372	0	118	64	23	89	0.0	1.731	21	neg
## 373	0	84	64	22	66	35.8	0.545	21	neg
## 374	2	105	58	40	94	34.9	0.225	25	neg
## 375	2	122	52	43	158	36.2	0.816	28	neg
## 376	12	140	82	43	325	39.2	0.528	58	pos
## 377	0	98	82	15	84	25.2	0.299	22	neg
## 378	1	87	60	37	75	37.2	0.509	22	neg
## 379	4	156	75	0	0	48.3	0.238	32	pos
## 380	0	93	100	39	72	43.4	1.021	35	neg
## 381	1	107	72	30	82	30.8	0.821	24	neg
## 382	0	105	68	22	0	20.0	0.236	22	neg
## 383	1	109	60	8	182	25.4	0.947	21	neg
## 384	1	90	62	18	59	25.1	1.268	25	neg
## 385	1	125	70	24	110	24.3	0.221	25	neg
## 386	1	119	54	13	50	22.3	0.205	24	neg
## 387	5	116	74	29	0	32.3	0.660	35	pos
## 388	8	105	100	36	0	43.3	0.239	45	pos
## 389	5	144	82	26	285	32.0	0.452	58	pos
## 390	3	100	68	23	81	31.6	0.949	28	neg
## 391	1	100	66	29	196	32.0	0.444	42	neg
## 392	5	166	76	0	0	45.7	0.340	27	pos
## 393	1	131	64	14	415	23.7	0.389	21	neg
## 394	4	116	72	12	87	22.1	0.463	37	neg
## 395	4	158	78	0	0	32.9	0.803	31	pos

## 396	2	127	58	24	275	27.7	1.600	25	neg
## 397	3	96	56	34	115	24.7	0.944	39	neg
## 398	0	131	66	40	0	34.3	0.196	22	pos
## 399	3	82	70	0	0	21.1	0.389	25	neg
## 400	3	193	70	31	0	34.9	0.241	25	pos
## 401	4	95	64	0	0	32.0	0.161	31	pos
## 402	6	137	61	0	0	24.2	0.151	55	neg
## 403	5	136	84	41	88	35.0	0.286	35	pos
## 404	9	72	78	25	0	31.6	0.280	38	neg
## 405	5	168	64	0	0	32.9	0.135	41	pos
## 406	2	123	48	32	165	42.1	0.520	26	neg
## 407	4	115	72	0	0	28.9	0.376	46	pos
## 408	0	101	62	0	0	21.9	0.336	25	neg
## 409	8	197	74	0	0	25.9	1.191	39	pos
## 410	1	172	68	49	579	42.4	0.702	28	pos
## 411	6	102	90	39	0	35.7	0.674	28	neg
## 412	1	112	72	30	176	34.4	0.528	25	neg
## 413	1	143	84	23	310	42.4	1.076	22	neg
## 414	1	143	74	22	61	26.2	0.256	21	neg
## 415	0	138	60	35	167	34.6	0.534	21	pos
## 416	3	173	84	33	474	35.7	0.258	22	pos
## 417	1	97	68	21	0	27.2	1.095	22	neg
## 418	4	144	82	32	0	38.5	0.554	37	pos
## 419	1	83	68	0	0	18.2	0.624	27	neg
## 420	3	129	64	29	115	26.4	0.219	28	pos
## 421	1	119	88	41	170	45.3	0.507	26	neg
## 422	2	94	68	18	76	26.0	0.561	21	neg
## 423	0	102	64	46	78	40.6	0.496	21	neg
## 424	2	115	64	22	0	30.8	0.421	21	neg
## 425	8	151	78	32	210	42.9	0.516	36	pos
## 426	4	184	78	39	277	37.0	0.264	31	pos
## 427	0	94	0	0	0	0.0	0.256	25	neg
## 428	1	181	64	30	180	34.1	0.328	38	pos
## 429	0	135	94	46	145	40.6	0.284	26	neg
## 430	1	95	82	25	180	35.0	0.233	43	pos
## 431	2	99	0	0	0	22.2	0.108	23	neg
## 432	3	89	74	16	85	30.4	0.551	38	neg
## 433	1	80	74	11	60	30.0	0.527	22	neg
## 434	2	139	75	0	0	25.6	0.167	29	neg
## 435	1	90	68	8	0	24.5	1.138	36	neg
## 436	0	141	0	0	0	42.4	0.205	29	pos
## 437	12	140	85	33	0	37.4	0.244	41	neg
## 438	5	147	75	0	0	29.9	0.434	28	neg
## 439	1	97	70	15	0	18.2	0.147	21	neg
## 440	6	107	88	0	0	36.8	0.727	31	neg
## 441	0	189	104	25	0	34.3	0.435	41	pos
## 442	2	83	66	23	50	32.2	0.497	22	neg
## 443	4	117	64	27	120	33.2	0.230	24	neg
## 444	8	108	70	0	0	30.5	0.955	33	pos
## 445	4	117	62	12	0	29.7	0.380	30	pos
## 446	0	180	78	63	14	59.4	2.420	25	pos
## 447	1	100	72	12	70	25.3	0.658	28	neg
## 448	0	95	80	45	92	36.5	0.330	26	neg
## 449	0	104	64	37	64	33.6	0.510	22	pos

## 450	0	120	74	18	63	30.5	0.285	26	neg
## 451	1	82	64	13	95	21.2	0.415	23	neg
## 452	2	134	70	0	0	28.9	0.542	23	pos
## 453	0	91	68	32	210	39.9	0.381	25	neg
## 454	2	119	0	0	0	19.6	0.832	72	neg
## 455	2	100	54	28	105	37.8	0.498	24	neg
## 456	14	175	62	30	0	33.6	0.212	38	pos
## 457	1	135	54	0	0	26.7	0.687	62	neg
## 458	5	86	68	28	71	30.2	0.364	24	neg
## 459	10	148	84	48	237	37.6	1.001	51	pos
## 460	9	134	74	33	60	25.9	0.460	81	neg
## 461	9	120	72	22	56	20.8	0.733	48	neg
## 462	1	71	62	0	0	21.8	0.416	26	neg
## 463	8	74	70	40	49	35.3	0.705	39	neg
## 464	5	88	78	30	0	27.6	0.258	37	neg
## 465	10	115	98	0	0	24.0	1.022	34	neg
## 466	0	124	56	13	105	21.8	0.452	21	neg
## 467	0	74	52	10	36	27.8	0.269	22	neg
## 468	0	97	64	36	100	36.8	0.600	25	neg
## 469	8	120	0	0	0	30.0	0.183	38	pos
## 470	6	154	78	41	140	46.1	0.571	27	neg
## 471	1	144	82	40	0	41.3	0.607	28	neg
## 472	0	137	70	38	0	33.2	0.170	22	neg
## 473	0	119	66	27	0	38.8	0.259	22	neg
## 474	7	136	90	0	0	29.9	0.210	50	neg
## 475	4	114	64	0	0	28.9	0.126	24	neg
## 476	0	137	84	27	0	27.3	0.231	59	neg
## 477	2	105	80	45	191	33.7	0.711	29	pos
## 478	7	114	76	17	110	23.8	0.466	31	neg
## 479	8	126	74	38	75	25.9	0.162	39	neg
## 480	4	132	86	31	0	28.0	0.419	63	neg
## 481	3	158	70	30	328	35.5	0.344	35	pos
## 482	0	123	88	37	0	35.2	0.197	29	neg
## 483	4	85	58	22	49	27.8	0.306	28	neg
## 484	0	84	82	31	125	38.2	0.233	23	neg
## 485	0	145	0	0	0	44.2	0.630	31	pos
## 486	0	135	68	42	250	42.3	0.365	24	pos
## 487	1	139	62	41	480	40.7	0.536	21	neg
## 488	0	173	78	32	265	46.5	1.159	58	neg
## 489	4	99	72	17	0	25.6	0.294	28	neg
## 490	8	194	80	0	0	26.1	0.551	67	neg
## 491	2	83	65	28	66	36.8	0.629	24	neg
## 492	2	89	90	30	0	33.5	0.292	42	neg
## 493	4	99	68	38	0	32.8	0.145	33	neg
## 494	4	125	70	18	122	28.9	1.144	45	pos
## 495	3	80	0	0	0	0.0	0.174	22	neg
## 496	6	166	74	0	0	26.6	0.304	66	neg
## 497	5	110	68	0	0	26.0	0.292	30	neg
## 498	2	81	72	15	76	30.1	0.547	25	neg
## 499	7	195	70	33	145	25.1	0.163	55	pos
## 500	6	154	74	32	193	29.3	0.839	39	neg
## 501	2	117	90	19	71	25.2	0.313	21	neg
## 502	3	84	72	32	0	37.2	0.267	28	neg
## 503	6	0	68	41	0	39.0	0.727	41	pos

## 504	7	94	64	25	79	33.3	0.738	41	neg
## 505	3	96	78	39	0	37.3	0.238	40	neg
## 506	10	75	82	0	0	33.3	0.263	38	neg
## 507	0	180	90	26	90	36.5	0.314	35	pos
## 508	1	130	60	23	170	28.6	0.692	21	neg
## 509	2	84	50	23	76	30.4	0.968	21	neg
## 510	8	120	78	0	0	25.0	0.409	64	neg
## 511	12	84	72	31	0	29.7	0.297	46	pos
## 512	0	139	62	17	210	22.1	0.207	21	neg
## 513	9	91	68	0	0	24.2	0.200	58	neg
## 514	2	91	62	0	0	27.3	0.525	22	neg
## 515	3	99	54	19	86	25.6	0.154	24	neg
## 516	3	163	70	18	105	31.6	0.268	28	pos
## 517	9	145	88	34	165	30.3	0.771	53	pos
## 518	7	125	86	0	0	37.6	0.304	51	neg
## 519	13	76	60	0	0	32.8	0.180	41	neg
## 520	6	129	90	7	326	19.6	0.582	60	neg
## 521	2	68	70	32	66	25.0	0.187	25	neg
## 522	3	124	80	33	130	33.2	0.305	26	neg
## 523	6	114	0	0	0	0.0	0.189	26	neg
## 524	9	130	70	0	0	34.2	0.652	45	pos
## 525	3	125	58	0	0	31.6	0.151	24	neg
## 526	3	87	60	18	0	21.8	0.444	21	neg
## 527	1	97	64	19	82	18.2	0.299	21	neg
## 528	3	116	74	15	105	26.3	0.107	24	neg
## 529	0	117	66	31	188	30.8	0.493	22	neg
## 530	0	111	65	0	0	24.6	0.660	31	neg
## 531	2	122	60	18	106	29.8	0.717	22	neg
## 532	0	107	76	0	0	45.3	0.686	24	neg
## 533	1	86	66	52	65	41.3	0.917	29	neg
## 534	6	91	0	0	0	29.8	0.501	31	neg
## 535	1	77	56	30	56	33.3	1.251	24	neg
## 536	4	132	0	0	0	32.9	0.302	23	pos
## 537	0	105	90	0	0	29.6	0.197	46	neg
## 538	0	57	60	0	0	21.7	0.735	67	neg
## 539	0	127	80	37	210	36.3	0.804	23	neg
## 540	3	129	92	49	155	36.4	0.968	32	pos
## 541	8	100	74	40	215	39.4	0.661	43	pos
## 542	3	128	72	25	190	32.4	0.549	27	pos
## 543	10	90	85	32	0	34.9	0.825	56	pos
## 544	4	84	90	23	56	39.5	0.159	25	neg
## 545	1	88	78	29	76	32.0	0.365	29	neg
## 546	8	186	90	35	225	34.5	0.423	37	pos
## 547	5	187	76	27	207	43.6	1.034	53	pos
## 548	4	131	68	21	166	33.1	0.160	28	neg
## 549	1	164	82	43	67	32.8	0.341	50	neg
## 550	4	189	110	31	0	28.5	0.680	37	neg
## 551	1	116	70	28	0	27.4	0.204	21	neg
## 552	3	84	68	30	106	31.9	0.591	25	neg
## 553	6	114	88	0	0	27.8	0.247	66	neg
## 554	1	88	62	24	44	29.9	0.422	23	neg
## 555	1	84	64	23	115	36.9	0.471	28	neg
## 556	7	124	70	33	215	25.5	0.161	37	neg
## 557	1	97	70	40	0	38.1	0.218	30	neg

## 558	8	110	76	0	0 27.8	0.237	58	neg
## 559	11	103	68	40	0 46.2	0.126	42	neg
## 560	11	85	74	0	0 30.1	0.300	35	neg
## 561	6	125	76	0	0 33.8	0.121	54	pos
## 562	0	198	66	32	274 41.3	0.502	28	pos
## 563	1	87	68	34	77 37.6	0.401	24	neg
## 564	6	99	60	19	54 26.9	0.497	32	neg
## 565	0	91	80	0	0 32.4	0.601	27	neg
## 566	2	95	54	14	88 26.1	0.748	22	neg
## 567	1	99	72	30	18 38.6	0.412	21	neg
## 568	6	92	62	32	126 32.0	0.085	46	neg
## 569	4	154	72	29	126 31.3	0.338	37	neg
## 570	0	121	66	30	165 34.3	0.203	33	pos
## 571	3	78	70	0	0 32.5	0.270	39	neg
## 572	2	130	96	0	0 22.6	0.268	21	neg
## 573	3	111	58	31	44 29.5	0.430	22	neg
## 574	2	98	60	17	120 34.7	0.198	22	neg
## 575	1	143	86	30	330 30.1	0.892	23	neg
## 576	1	119	44	47	63 35.5	0.280	25	neg
## 577	6	108	44	20	130 24.0	0.813	35	neg
## 578	2	118	80	0	0 42.9	0.693	21	pos
## 579	10	133	68	0	0 27.0	0.245	36	neg
## 580	2	197	70	99	0 34.7	0.575	62	pos
## 581	0	151	90	46	0 42.1	0.371	21	pos
## 582	6	109	60	27	0 25.0	0.206	27	neg
## 583	12	121	78	17	0 26.5	0.259	62	neg
## 584	8	100	76	0	0 38.7	0.190	42	neg
## 585	8	124	76	24	600 28.7	0.687	52	pos
## 586	1	93	56	11	0 22.5	0.417	22	neg
## 587	8	143	66	0	0 34.9	0.129	41	pos
## 588	6	103	66	0	0 24.3	0.249	29	neg
## 589	3	176	86	27	156 33.3	1.154	52	pos
## 590	0	73	0	0	0 21.1	0.342	25	neg
## 591	11	111	84	40	0 46.8	0.925	45	pos
## 592	2	112	78	50	140 39.4	0.175	24	neg
## 593	3	132	80	0	0 34.4	0.402	44	pos
## 594	2	82	52	22	115 28.5	1.699	25	neg
## 595	6	123	72	45	230 33.6	0.733	34	neg
## 596	0	188	82	14	185 32.0	0.682	22	pos
## 597	0	67	76	0	0 45.3	0.194	46	neg
## 598	1	89	24	19	25 27.8	0.559	21	neg
## 599	1	173	74	0	0 36.8	0.088	38	pos
## 600	1	109	38	18	120 23.1	0.407	26	neg
## 601	1	108	88	19	0 27.1	0.400	24	neg
## 602	6	96	0	0	0 23.7	0.190	28	neg
## 603	1	124	74	36	0 27.8	0.100	30	neg
## 604	7	150	78	29	126 35.2	0.692	54	pos
## 605	4	183	0	0	0 28.4	0.212	36	pos
## 606	1	124	60	32	0 35.8	0.514	21	neg
## 607	1	181	78	42	293 40.0	1.258	22	pos
## 608	1	92	62	25	41 19.5	0.482	25	neg
## 609	0	152	82	39	272 41.5	0.270	27	neg
## 610	1	111	62	13	182 24.0	0.138	23	neg
## 611	3	106	54	21	158 30.9	0.292	24	neg

## 612	3	174	58	22	194	32.9	0.593	36	pos
## 613	7	168	88	42	321	38.2	0.787	40	pos
## 614	6	105	80	28	0	32.5	0.878	26	neg
## 615	11	138	74	26	144	36.1	0.557	50	pos
## 616	3	106	72	0	0	25.8	0.207	27	neg
## 617	6	117	96	0	0	28.7	0.157	30	neg
## 618	2	68	62	13	15	20.1	0.257	23	neg
## 619	9	112	82	24	0	28.2	1.282	50	pos
## 620	0	119	0	0	0	32.4	0.141	24	pos
## 621	2	112	86	42	160	38.4	0.246	28	neg
## 622	2	92	76	20	0	24.2	1.698	28	neg
## 623	6	183	94	0	0	40.8	1.461	45	neg
## 624	0	94	70	27	115	43.5	0.347	21	neg
## 625	2	108	64	0	0	30.8	0.158	21	neg
## 626	4	90	88	47	54	37.7	0.362	29	neg
## 627	0	125	68	0	0	24.7	0.206	21	neg
## 628	0	132	78	0	0	32.4	0.393	21	neg
## 629	5	128	80	0	0	34.6	0.144	45	neg
## 630	4	94	65	22	0	24.7	0.148	21	neg
## 631	7	114	64	0	0	27.4	0.732	34	pos
## 632	0	102	78	40	90	34.5	0.238	24	neg
## 633	2	111	60	0	0	26.2	0.343	23	neg
## 634	1	128	82	17	183	27.5	0.115	22	neg
## 635	10	92	62	0	0	25.9	0.167	31	neg
## 636	13	104	72	0	0	31.2	0.465	38	pos
## 637	5	104	74	0	0	28.8	0.153	48	neg
## 638	2	94	76	18	66	31.6	0.649	23	neg
## 639	7	97	76	32	91	40.9	0.871	32	pos
## 640	1	100	74	12	46	19.5	0.149	28	neg
## 641	0	102	86	17	105	29.3	0.695	27	neg
## 642	4	128	70	0	0	34.3	0.303	24	neg
## 643	6	147	80	0	0	29.5	0.178	50	pos
## 644	4	90	0	0	0	28.0	0.610	31	neg
## 645	3	103	72	30	152	27.6	0.730	27	neg
## 646	2	157	74	35	440	39.4	0.134	30	neg
## 647	1	167	74	17	144	23.4	0.447	33	pos
## 648	0	179	50	36	159	37.8	0.455	22	pos
## 649	11	136	84	35	130	28.3	0.260	42	pos
## 650	0	107	60	25	0	26.4	0.133	23	neg
## 651	1	91	54	25	100	25.2	0.234	23	neg
## 652	1	117	60	23	106	33.8	0.466	27	neg
## 653	5	123	74	40	77	34.1	0.269	28	neg
## 654	2	120	54	0	0	26.8	0.455	27	neg
## 655	1	106	70	28	135	34.2	0.142	22	neg
## 656	2	155	52	27	540	38.7	0.240	25	pos
## 657	2	101	58	35	90	21.8	0.155	22	neg
## 658	1	120	80	48	200	38.9	1.162	41	neg
## 659	11	127	106	0	0	39.0	0.190	51	neg
## 660	3	80	82	31	70	34.2	1.292	27	pos
## 661	10	162	84	0	0	27.7	0.182	54	neg
## 662	1	199	76	43	0	42.9	1.394	22	pos
## 663	8	167	106	46	231	37.6	0.165	43	pos
## 664	9	145	80	46	130	37.9	0.637	40	pos
## 665	6	115	60	39	0	33.7	0.245	40	pos

## 666	1	112	80	45	132	34.8	0.217	24	neg
## 667	4	145	82	18	0	32.5	0.235	70	pos
## 668	10	111	70	27	0	27.5	0.141	40	pos
## 669	6	98	58	33	190	34.0	0.430	43	neg
## 670	9	154	78	30	100	30.9	0.164	45	neg
## 671	6	165	68	26	168	33.6	0.631	49	neg
## 672	1	99	58	10	0	25.4	0.551	21	neg
## 673	10	68	106	23	49	35.5	0.285	47	neg
## 674	3	123	100	35	240	57.3	0.880	22	neg
## 675	8	91	82	0	0	35.6	0.587	68	neg
## 676	6	195	70	0	0	30.9	0.328	31	pos
## 677	9	156	86	0	0	24.8	0.230	53	pos
## 678	0	93	60	0	0	35.3	0.263	25	neg
## 679	3	121	52	0	0	36.0	0.127	25	pos
## 680	2	101	58	17	265	24.2	0.614	23	neg
## 681	2	56	56	28	45	24.2	0.332	22	neg
## 682	0	162	76	36	0	49.6	0.364	26	pos
## 683	0	95	64	39	105	44.6	0.366	22	neg
## 684	4	125	80	0	0	32.3	0.536	27	pos
## 685	5	136	82	0	0	0.0	0.640	69	neg
## 686	2	129	74	26	205	33.2	0.591	25	neg
## 687	3	130	64	0	0	23.1	0.314	22	neg
## 688	1	107	50	19	0	28.3	0.181	29	neg
## 689	1	140	74	26	180	24.1	0.828	23	neg
## 690	1	144	82	46	180	46.1	0.335	46	pos
## 691	8	107	80	0	0	24.6	0.856	34	neg
## 692	13	158	114	0	0	42.3	0.257	44	pos
## 693	2	121	70	32	95	39.1	0.886	23	neg
## 694	7	129	68	49	125	38.5	0.439	43	pos
## 695	2	90	60	0	0	23.5	0.191	25	neg
## 696	7	142	90	24	480	30.4	0.128	43	pos
## 697	3	169	74	19	125	29.9	0.268	31	pos
## 698	0	99	0	0	0	25.0	0.253	22	neg
## 699	4	127	88	11	155	34.5	0.598	28	neg
## 700	4	118	70	0	0	44.5	0.904	26	neg
## 701	2	122	76	27	200	35.9	0.483	26	neg
## 702	6	125	78	31	0	27.6	0.565	49	pos
## 703	1	168	88	29	0	35.0	0.905	52	pos
## 704	2	129	0	0	0	38.5	0.304	41	neg
## 705	4	110	76	20	100	28.4	0.118	27	neg
## 706	6	80	80	36	0	39.8	0.177	28	neg
## 707	10	115	0	0	0	0.0	0.261	30	pos
## 708	2	127	46	21	335	34.4	0.176	22	neg
## 709	9	164	78	0	0	32.8	0.148	45	pos
## 710	2	93	64	32	160	38.0	0.674	23	pos
## 711	3	158	64	13	387	31.2	0.295	24	neg
## 712	5	126	78	27	22	29.6	0.439	40	neg
## 713	10	129	62	36	0	41.2	0.441	38	pos
## 714	0	134	58	20	291	26.4	0.352	21	neg
## 715	3	102	74	0	0	29.5	0.121	32	neg
## 716	7	187	50	33	392	33.9	0.826	34	pos
## 717	3	173	78	39	185	33.8	0.970	31	pos
## 718	10	94	72	18	0	23.1	0.595	56	neg
## 719	1	108	60	46	178	35.5	0.415	24	neg

```
## 720      5      97      76      27      0 35.6      0.378 52      pos
## 721      4      83      86      19      0 29.3      0.317 34      neg
## 722      1     114      66      36     200 38.1      0.289 21      neg
## 723      1     149      68      29     127 29.3      0.349 42      pos
## 724      5     117      86      30     105 39.1      0.251 42      neg
## 725      1     111      94       0      0 32.8      0.265 45      neg
## 726      4     112      78      40      0 39.4      0.236 38      neg
## 727      1     116      78      29     180 36.1      0.496 25      neg
## 728      0     141      84      26      0 32.4      0.433 22      neg
## 729      2     175      88       0      0 22.9      0.326 22      neg
## 730      2      92      52       0      0 30.1      0.141 22      neg
## 731      3     130      78      23      79 28.4      0.323 34      pos
## 732      8     120      86       0      0 28.4      0.259 22      pos
## 733      2     174      88      37     120 44.5      0.646 24      pos
## 734      2     106      56      27     165 29.0      0.426 22      neg
## 735      2     105      75       0      0 23.3      0.560 53      neg
## 736      4      95      60      32      0 35.4      0.284 28      neg
## 737      0     126      86      27     120 27.4      0.515 21      neg
## 738      8      65      72      23      0 32.0      0.600 42      neg
## 739      2      99      60      17     160 36.6      0.453 21      neg
## 740      1     102      74       0      0 39.5      0.293 42      pos
## 741     11     120      80      37     150 42.3      0.785 48      pos
## 742      3     102      44      20      94 30.8      0.400 26      neg
## 743      1     109      58      18     116 28.5      0.219 22      neg
## 744      9     140      94       0      0 32.7      0.734 45      pos
## 745     13     153      88      37     140 40.6      1.174 39      neg
## 746     12     100      84      33     105 30.0      0.488 46      neg
## 747      1     147      94      41      0 49.3      0.358 27      pos
## 748      1      81      74      41      57 46.3      1.096 32      neg
## 749      3     187      70      22     200 36.4      0.408 36      pos
## 750      6     162      62       0      0 24.3      0.178 50      pos
## 751      4     136      70       0      0 31.2      1.182 22      pos
## 752      1     121      78      39      74 39.0      0.261 28      neg
## 753      3     108      62      24      0 26.0      0.223 25      neg
## 754      0     181      88      44     510 43.3      0.222 26      pos
## 755      8     154      78      32      0 32.4      0.443 45      pos
## 756      1     128      88      39     110 36.5      1.057 37      pos
## 757      7     137      90      41      0 32.0      0.391 39      neg
## 758      0     123      72       0      0 36.3      0.258 52      pos
## 759      1     106      76       0      0 37.5      0.197 26      neg
## 760      6     190      92       0      0 35.5      0.278 66      pos
## 761      2      88      58      26      16 28.4      0.766 22      neg
## 762      9     170      74      31      0 44.0      0.403 43      pos
## 763      9      89      62       0      0 22.5      0.142 33      neg
## 764     10     101      76      48     180 32.9      0.171 63      neg
## 765      2     122      70      27      0 36.8      0.340 27      neg
## 766      5     121      72      23     112 26.2      0.245 30      neg
## 767      1     126      60       0      0 30.1      0.349 47      pos
## 768      1      93      70      31      0 30.4      0.315 23      neg
```

```
#make diabetes numeric by making the pos and neg values binary
PimaIndiansDiabetes$diabetes <- ifelse(PimaIndiansDiabetes$diabetes == "pos", 1, 0)
PimaIndiansDiabetes
```

```
##      pregnant glucose pressure triceps insulin mass pedigree age diabetes
```

## 1	6	148	72	35	0	33.6	0.627	50	1
## 2	1	85	66	29	0	26.6	0.351	31	0
## 3	8	183	64	0	0	23.3	0.672	32	1
## 4	1	89	66	23	94	28.1	0.167	21	0
## 5	0	137	40	35	168	43.1	2.288	33	1
## 6	5	116	74	0	0	25.6	0.201	30	0
## 7	3	78	50	32	88	31.0	0.248	26	1
## 8	10	115	0	0	0	35.3	0.134	29	0
## 9	2	197	70	45	543	30.5	0.158	53	1
## 10	8	125	96	0	0	0.0	0.232	54	1
## 11	4	110	92	0	0	37.6	0.191	30	0
## 12	10	168	74	0	0	38.0	0.537	34	1
## 13	10	139	80	0	0	27.1	1.441	57	0
## 14	1	189	60	23	846	30.1	0.398	59	1
## 15	5	166	72	19	175	25.8	0.587	51	1
## 16	7	100	0	0	0	30.0	0.484	32	1
## 17	0	118	84	47	230	45.8	0.551	31	1
## 18	7	107	74	0	0	29.6	0.254	31	1
## 19	1	103	30	38	83	43.3	0.183	33	0
## 20	1	115	70	30	96	34.6	0.529	32	1
## 21	3	126	88	41	235	39.3	0.704	27	0
## 22	8	99	84	0	0	35.4	0.388	50	0
## 23	7	196	90	0	0	39.8	0.451	41	1
## 24	9	119	80	35	0	29.0	0.263	29	1
## 25	11	143	94	33	146	36.6	0.254	51	1
## 26	10	125	70	26	115	31.1	0.205	41	1
## 27	7	147	76	0	0	39.4	0.257	43	1
## 28	1	97	66	15	140	23.2	0.487	22	0
## 29	13	145	82	19	110	22.2	0.245	57	0
## 30	5	117	92	0	0	34.1	0.337	38	0
## 31	5	109	75	26	0	36.0	0.546	60	0
## 32	3	158	76	36	245	31.6	0.851	28	1
## 33	3	88	58	11	54	24.8	0.267	22	0
## 34	6	92	92	0	0	19.9	0.188	28	0
## 35	10	122	78	31	0	27.6	0.512	45	0
## 36	4	103	60	33	192	24.0	0.966	33	0
## 37	11	138	76	0	0	33.2	0.420	35	0
## 38	9	102	76	37	0	32.9	0.665	46	1
## 39	2	90	68	42	0	38.2	0.503	27	1
## 40	4	111	72	47	207	37.1	1.390	56	1
## 41	3	180	64	25	70	34.0	0.271	26	0
## 42	7	133	84	0	0	40.2	0.696	37	0
## 43	7	106	92	18	0	22.7	0.235	48	0
## 44	9	171	110	24	240	45.4	0.721	54	1
## 45	7	159	64	0	0	27.4	0.294	40	0
## 46	0	180	66	39	0	42.0	1.893	25	1
## 47	1	146	56	0	0	29.7	0.564	29	0
## 48	2	71	70	27	0	28.0	0.586	22	0
## 49	7	103	66	32	0	39.1	0.344	31	1
## 50	7	105	0	0	0	0.0	0.305	24	0
## 51	1	103	80	11	82	19.4	0.491	22	0
## 52	1	101	50	15	36	24.2	0.526	26	0
## 53	5	88	66	21	23	24.4	0.342	30	0
## 54	8	176	90	34	300	33.7	0.467	58	1

## 55	7	150	66	42	342	34.7	0.718	42	0
## 56	1	73	50	10	0	23.0	0.248	21	0
## 57	7	187	68	39	304	37.7	0.254	41	1
## 58	0	100	88	60	110	46.8	0.962	31	0
## 59	0	146	82	0	0	40.5	1.781	44	0
## 60	0	105	64	41	142	41.5	0.173	22	0
## 61	2	84	0	0	0	0.0	0.304	21	0
## 62	8	133	72	0	0	32.9	0.270	39	1
## 63	5	44	62	0	0	25.0	0.587	36	0
## 64	2	141	58	34	128	25.4	0.699	24	0
## 65	7	114	66	0	0	32.8	0.258	42	1
## 66	5	99	74	27	0	29.0	0.203	32	0
## 67	0	109	88	30	0	32.5	0.855	38	1
## 68	2	109	92	0	0	42.7	0.845	54	0
## 69	1	95	66	13	38	19.6	0.334	25	0
## 70	4	146	85	27	100	28.9	0.189	27	0
## 71	2	100	66	20	90	32.9	0.867	28	1
## 72	5	139	64	35	140	28.6	0.411	26	0
## 73	13	126	90	0	0	43.4	0.583	42	1
## 74	4	129	86	20	270	35.1	0.231	23	0
## 75	1	79	75	30	0	32.0	0.396	22	0
## 76	1	0	48	20	0	24.7	0.140	22	0
## 77	7	62	78	0	0	32.6	0.391	41	0
## 78	5	95	72	33	0	37.7	0.370	27	0
## 79	0	131	0	0	0	43.2	0.270	26	1
## 80	2	112	66	22	0	25.0	0.307	24	0
## 81	3	113	44	13	0	22.4	0.140	22	0
## 82	2	74	0	0	0	0.0	0.102	22	0
## 83	7	83	78	26	71	29.3	0.767	36	0
## 84	0	101	65	28	0	24.6	0.237	22	0
## 85	5	137	108	0	0	48.8	0.227	37	1
## 86	2	110	74	29	125	32.4	0.698	27	0
## 87	13	106	72	54	0	36.6	0.178	45	0
## 88	2	100	68	25	71	38.5	0.324	26	0
## 89	15	136	70	32	110	37.1	0.153	43	1
## 90	1	107	68	19	0	26.5	0.165	24	0
## 91	1	80	55	0	0	19.1	0.258	21	0
## 92	4	123	80	15	176	32.0	0.443	34	0
## 93	7	81	78	40	48	46.7	0.261	42	0
## 94	4	134	72	0	0	23.8	0.277	60	1
## 95	2	142	82	18	64	24.7	0.761	21	0
## 96	6	144	72	27	228	33.9	0.255	40	0
## 97	2	92	62	28	0	31.6	0.130	24	0
## 98	1	71	48	18	76	20.4	0.323	22	0
## 99	6	93	50	30	64	28.7	0.356	23	0
## 100	1	122	90	51	220	49.7	0.325	31	1
## 101	1	163	72	0	0	39.0	1.222	33	1
## 102	1	151	60	0	0	26.1	0.179	22	0
## 103	0	125	96	0	0	22.5	0.262	21	0
## 104	1	81	72	18	40	26.6	0.283	24	0
## 105	2	85	65	0	0	39.6	0.930	27	0
## 106	1	126	56	29	152	28.7	0.801	21	0
## 107	1	96	122	0	0	22.4	0.207	27	0
## 108	4	144	58	28	140	29.5	0.287	37	0

## 109	3	83	58	31	18 34.3	0.336	25	0
## 110	0	95	85	25	36 37.4	0.247	24	1
## 111	3	171	72	33	135 33.3	0.199	24	1
## 112	8	155	62	26	495 34.0	0.543	46	1
## 113	1	89	76	34	37 31.2	0.192	23	0
## 114	4	76	62	0	0 34.0	0.391	25	0
## 115	7	160	54	32	175 30.5	0.588	39	1
## 116	4	146	92	0	0 31.2	0.539	61	1
## 117	5	124	74	0	0 34.0	0.220	38	1
## 118	5	78	48	0	0 33.7	0.654	25	0
## 119	4	97	60	23	0 28.2	0.443	22	0
## 120	4	99	76	15	51 23.2	0.223	21	0
## 121	0	162	76	56	100 53.2	0.759	25	1
## 122	6	111	64	39	0 34.2	0.260	24	0
## 123	2	107	74	30	100 33.6	0.404	23	0
## 124	5	132	80	0	0 26.8	0.186	69	0
## 125	0	113	76	0	0 33.3	0.278	23	1
## 126	1	88	30	42	99 55.0	0.496	26	1
## 127	3	120	70	30	135 42.9	0.452	30	0
## 128	1	118	58	36	94 33.3	0.261	23	0
## 129	1	117	88	24	145 34.5	0.403	40	1
## 130	0	105	84	0	0 27.9	0.741	62	1
## 131	4	173	70	14	168 29.7	0.361	33	1
## 132	9	122	56	0	0 33.3	1.114	33	1
## 133	3	170	64	37	225 34.5	0.356	30	1
## 134	8	84	74	31	0 38.3	0.457	39	0
## 135	2	96	68	13	49 21.1	0.647	26	0
## 136	2	125	60	20	140 33.8	0.088	31	0
## 137	0	100	70	26	50 30.8	0.597	21	0
## 138	0	93	60	25	92 28.7	0.532	22	0
## 139	0	129	80	0	0 31.2	0.703	29	0
## 140	5	105	72	29	325 36.9	0.159	28	0
## 141	3	128	78	0	0 21.1	0.268	55	0
## 142	5	106	82	30	0 39.5	0.286	38	0
## 143	2	108	52	26	63 32.5	0.318	22	0
## 144	10	108	66	0	0 32.4	0.272	42	1
## 145	4	154	62	31	284 32.8	0.237	23	0
## 146	0	102	75	23	0 0.0	0.572	21	0
## 147	9	57	80	37	0 32.8	0.096	41	0
## 148	2	106	64	35	119 30.5	1.400	34	0
## 149	5	147	78	0	0 33.7	0.218	65	0
## 150	2	90	70	17	0 27.3	0.085	22	0
## 151	1	136	74	50	204 37.4	0.399	24	0
## 152	4	114	65	0	0 21.9	0.432	37	0
## 153	9	156	86	28	155 34.3	1.189	42	1
## 154	1	153	82	42	485 40.6	0.687	23	0
## 155	8	188	78	0	0 47.9	0.137	43	1
## 156	7	152	88	44	0 50.0	0.337	36	1
## 157	2	99	52	15	94 24.6	0.637	21	0
## 158	1	109	56	21	135 25.2	0.833	23	0
## 159	2	88	74	19	53 29.0	0.229	22	0
## 160	17	163	72	41	114 40.9	0.817	47	1
## 161	4	151	90	38	0 29.7	0.294	36	0
## 162	7	102	74	40	105 37.2	0.204	45	0

## 163	0	114	80	34	285	44.2	0.167	27	0
## 164	2	100	64	23	0	29.7	0.368	21	0
## 165	0	131	88	0	0	31.6	0.743	32	1
## 166	6	104	74	18	156	29.9	0.722	41	1
## 167	3	148	66	25	0	32.5	0.256	22	0
## 168	4	120	68	0	0	29.6	0.709	34	0
## 169	4	110	66	0	0	31.9	0.471	29	0
## 170	3	111	90	12	78	28.4	0.495	29	0
## 171	6	102	82	0	0	30.8	0.180	36	1
## 172	6	134	70	23	130	35.4	0.542	29	1
## 173	2	87	0	23	0	28.9	0.773	25	0
## 174	1	79	60	42	48	43.5	0.678	23	0
## 175	2	75	64	24	55	29.7	0.370	33	0
## 176	8	179	72	42	130	32.7	0.719	36	1
## 177	6	85	78	0	0	31.2	0.382	42	0
## 178	0	129	110	46	130	67.1	0.319	26	1
## 179	5	143	78	0	0	45.0	0.190	47	0
## 180	5	130	82	0	0	39.1	0.956	37	1
## 181	6	87	80	0	0	23.2	0.084	32	0
## 182	0	119	64	18	92	34.9	0.725	23	0
## 183	1	0	74	20	23	27.7	0.299	21	0
## 184	5	73	60	0	0	26.8	0.268	27	0
## 185	4	141	74	0	0	27.6	0.244	40	0
## 186	7	194	68	28	0	35.9	0.745	41	1
## 187	8	181	68	36	495	30.1	0.615	60	1
## 188	1	128	98	41	58	32.0	1.321	33	1
## 189	8	109	76	39	114	27.9	0.640	31	1
## 190	5	139	80	35	160	31.6	0.361	25	1
## 191	3	111	62	0	0	22.6	0.142	21	0
## 192	9	123	70	44	94	33.1	0.374	40	0
## 193	7	159	66	0	0	30.4	0.383	36	1
## 194	11	135	0	0	0	52.3	0.578	40	1
## 195	8	85	55	20	0	24.4	0.136	42	0
## 196	5	158	84	41	210	39.4	0.395	29	1
## 197	1	105	58	0	0	24.3	0.187	21	0
## 198	3	107	62	13	48	22.9	0.678	23	1
## 199	4	109	64	44	99	34.8	0.905	26	1
## 200	4	148	60	27	318	30.9	0.150	29	1
## 201	0	113	80	16	0	31.0	0.874	21	0
## 202	1	138	82	0	0	40.1	0.236	28	0
## 203	0	108	68	20	0	27.3	0.787	32	0
## 204	2	99	70	16	44	20.4	0.235	27	0
## 205	6	103	72	32	190	37.7	0.324	55	0
## 206	5	111	72	28	0	23.9	0.407	27	0
## 207	8	196	76	29	280	37.5	0.605	57	1
## 208	5	162	104	0	0	37.7	0.151	52	1
## 209	1	96	64	27	87	33.2	0.289	21	0
## 210	7	184	84	33	0	35.5	0.355	41	1
## 211	2	81	60	22	0	27.7	0.290	25	0
## 212	0	147	85	54	0	42.8	0.375	24	0
## 213	7	179	95	31	0	34.2	0.164	60	0
## 214	0	140	65	26	130	42.6	0.431	24	1
## 215	9	112	82	32	175	34.2	0.260	36	1
## 216	12	151	70	40	271	41.8	0.742	38	1

## 217	5	109	62	41	129	35.8	0.514	25	1
## 218	6	125	68	30	120	30.0	0.464	32	0
## 219	5	85	74	22	0	29.0	1.224	32	1
## 220	5	112	66	0	0	37.8	0.261	41	1
## 221	0	177	60	29	478	34.6	1.072	21	1
## 222	2	158	90	0	0	31.6	0.805	66	1
## 223	7	119	0	0	0	25.2	0.209	37	0
## 224	7	142	60	33	190	28.8	0.687	61	0
## 225	1	100	66	15	56	23.6	0.666	26	0
## 226	1	87	78	27	32	34.6	0.101	22	0
## 227	0	101	76	0	0	35.7	0.198	26	0
## 228	3	162	52	38	0	37.2	0.652	24	1
## 229	4	197	70	39	744	36.7	2.329	31	0
## 230	0	117	80	31	53	45.2	0.089	24	0
## 231	4	142	86	0	0	44.0	0.645	22	1
## 232	6	134	80	37	370	46.2	0.238	46	1
## 233	1	79	80	25	37	25.4	0.583	22	0
## 234	4	122	68	0	0	35.0	0.394	29	0
## 235	3	74	68	28	45	29.7	0.293	23	0
## 236	4	171	72	0	0	43.6	0.479	26	1
## 237	7	181	84	21	192	35.9	0.586	51	1
## 238	0	179	90	27	0	44.1	0.686	23	1
## 239	9	164	84	21	0	30.8	0.831	32	1
## 240	0	104	76	0	0	18.4	0.582	27	0
## 241	1	91	64	24	0	29.2	0.192	21	0
## 242	4	91	70	32	88	33.1	0.446	22	0
## 243	3	139	54	0	0	25.6	0.402	22	1
## 244	6	119	50	22	176	27.1	1.318	33	1
## 245	2	146	76	35	194	38.2	0.329	29	0
## 246	9	184	85	15	0	30.0	1.213	49	1
## 247	10	122	68	0	0	31.2	0.258	41	0
## 248	0	165	90	33	680	52.3	0.427	23	0
## 249	9	124	70	33	402	35.4	0.282	34	0
## 250	1	111	86	19	0	30.1	0.143	23	0
## 251	9	106	52	0	0	31.2	0.380	42	0
## 252	2	129	84	0	0	28.0	0.284	27	0
## 253	2	90	80	14	55	24.4	0.249	24	0
## 254	0	86	68	32	0	35.8	0.238	25	0
## 255	12	92	62	7	258	27.6	0.926	44	1
## 256	1	113	64	35	0	33.6	0.543	21	1
## 257	3	111	56	39	0	30.1	0.557	30	0
## 258	2	114	68	22	0	28.7	0.092	25	0
## 259	1	193	50	16	375	25.9	0.655	24	0
## 260	11	155	76	28	150	33.3	1.353	51	1
## 261	3	191	68	15	130	30.9	0.299	34	0
## 262	3	141	0	0	0	30.0	0.761	27	1
## 263	4	95	70	32	0	32.1	0.612	24	0
## 264	3	142	80	15	0	32.4	0.200	63	0
## 265	4	123	62	0	0	32.0	0.226	35	1
## 266	5	96	74	18	67	33.6	0.997	43	0
## 267	0	138	0	0	0	36.3	0.933	25	1
## 268	2	128	64	42	0	40.0	1.101	24	0
## 269	0	102	52	0	0	25.1	0.078	21	0
## 270	2	146	0	0	0	27.5	0.240	28	1

## 271	10	101	86	37	0 45.6	1.136	38	1
## 272	2	108	62	32	56 25.2	0.128	21	0
## 273	3	122	78	0	0 23.0	0.254	40	0
## 274	1	71	78	50	45 33.2	0.422	21	0
## 275	13	106	70	0	0 34.2	0.251	52	0
## 276	2	100	70	52	57 40.5	0.677	25	0
## 277	7	106	60	24	0 26.5	0.296	29	1
## 278	0	104	64	23	116 27.8	0.454	23	0
## 279	5	114	74	0	0 24.9	0.744	57	0
## 280	2	108	62	10	278 25.3	0.881	22	0
## 281	0	146	70	0	0 37.9	0.334	28	1
## 282	10	129	76	28	122 35.9	0.280	39	0
## 283	7	133	88	15	155 32.4	0.262	37	0
## 284	7	161	86	0	0 30.4	0.165	47	1
## 285	2	108	80	0	0 27.0	0.259	52	1
## 286	7	136	74	26	135 26.0	0.647	51	0
## 287	5	155	84	44	545 38.7	0.619	34	0
## 288	1	119	86	39	220 45.6	0.808	29	1
## 289	4	96	56	17	49 20.8	0.340	26	0
## 290	5	108	72	43	75 36.1	0.263	33	0
## 291	0	78	88	29	40 36.9	0.434	21	0
## 292	0	107	62	30	74 36.6	0.757	25	1
## 293	2	128	78	37	182 43.3	1.224	31	1
## 294	1	128	48	45	194 40.5	0.613	24	1
## 295	0	161	50	0	0 21.9	0.254	65	0
## 296	6	151	62	31	120 35.5	0.692	28	0
## 297	2	146	70	38	360 28.0	0.337	29	1
## 298	0	126	84	29	215 30.7	0.520	24	0
## 299	14	100	78	25	184 36.6	0.412	46	1
## 300	8	112	72	0	0 23.6	0.840	58	0
## 301	0	167	0	0	0 32.3	0.839	30	1
## 302	2	144	58	33	135 31.6	0.422	25	1
## 303	5	77	82	41	42 35.8	0.156	35	0
## 304	5	115	98	0	0 52.9	0.209	28	1
## 305	3	150	76	0	0 21.0	0.207	37	0
## 306	2	120	76	37	105 39.7	0.215	29	0
## 307	10	161	68	23	132 25.5	0.326	47	1
## 308	0	137	68	14	148 24.8	0.143	21	0
## 309	0	128	68	19	180 30.5	1.391	25	1
## 310	2	124	68	28	205 32.9	0.875	30	1
## 311	6	80	66	30	0 26.2	0.313	41	0
## 312	0	106	70	37	148 39.4	0.605	22	0
## 313	2	155	74	17	96 26.6	0.433	27	1
## 314	3	113	50	10	85 29.5	0.626	25	0
## 315	7	109	80	31	0 35.9	1.127	43	1
## 316	2	112	68	22	94 34.1	0.315	26	0
## 317	3	99	80	11	64 19.3	0.284	30	0
## 318	3	182	74	0	0 30.5	0.345	29	1
## 319	3	115	66	39	140 38.1	0.150	28	0
## 320	6	194	78	0	0 23.5	0.129	59	1
## 321	4	129	60	12	231 27.5	0.527	31	0
## 322	3	112	74	30	0 31.6	0.197	25	1
## 323	0	124	70	20	0 27.4	0.254	36	1
## 324	13	152	90	33	29 26.8	0.731	43	1

## 325	2	112	75	32	0	35.7	0.148	21	0
## 326	1	157	72	21	168	25.6	0.123	24	0
## 327	1	122	64	32	156	35.1	0.692	30	1
## 328	10	179	70	0	0	35.1	0.200	37	0
## 329	2	102	86	36	120	45.5	0.127	23	1
## 330	6	105	70	32	68	30.8	0.122	37	0
## 331	8	118	72	19	0	23.1	1.476	46	0
## 332	2	87	58	16	52	32.7	0.166	25	0
## 333	1	180	0	0	0	43.3	0.282	41	1
## 334	12	106	80	0	0	23.6	0.137	44	0
## 335	1	95	60	18	58	23.9	0.260	22	0
## 336	0	165	76	43	255	47.9	0.259	26	0
## 337	0	117	0	0	0	33.8	0.932	44	0
## 338	5	115	76	0	0	31.2	0.343	44	1
## 339	9	152	78	34	171	34.2	0.893	33	1
## 340	7	178	84	0	0	39.9	0.331	41	1
## 341	1	130	70	13	105	25.9	0.472	22	0
## 342	1	95	74	21	73	25.9	0.673	36	0
## 343	1	0	68	35	0	32.0	0.389	22	0
## 344	5	122	86	0	0	34.7	0.290	33	0
## 345	8	95	72	0	0	36.8	0.485	57	0
## 346	8	126	88	36	108	38.5	0.349	49	0
## 347	1	139	46	19	83	28.7	0.654	22	0
## 348	3	116	0	0	0	23.5	0.187	23	0
## 349	3	99	62	19	74	21.8	0.279	26	0
## 350	5	0	80	32	0	41.0	0.346	37	1
## 351	4	92	80	0	0	42.2	0.237	29	0
## 352	4	137	84	0	0	31.2	0.252	30	0
## 353	3	61	82	28	0	34.4	0.243	46	0
## 354	1	90	62	12	43	27.2	0.580	24	0
## 355	3	90	78	0	0	42.7	0.559	21	0
## 356	9	165	88	0	0	30.4	0.302	49	1
## 357	1	125	50	40	167	33.3	0.962	28	1
## 358	13	129	0	30	0	39.9	0.569	44	1
## 359	12	88	74	40	54	35.3	0.378	48	0
## 360	1	196	76	36	249	36.5	0.875	29	1
## 361	5	189	64	33	325	31.2	0.583	29	1
## 362	5	158	70	0	0	29.8	0.207	63	0
## 363	5	103	108	37	0	39.2	0.305	65	0
## 364	4	146	78	0	0	38.5	0.520	67	1
## 365	4	147	74	25	293	34.9	0.385	30	0
## 366	5	99	54	28	83	34.0	0.499	30	0
## 367	6	124	72	0	0	27.6	0.368	29	1
## 368	0	101	64	17	0	21.0	0.252	21	0
## 369	3	81	86	16	66	27.5	0.306	22	0
## 370	1	133	102	28	140	32.8	0.234	45	1
## 371	3	173	82	48	465	38.4	2.137	25	1
## 372	0	118	64	23	89	0.0	1.731	21	0
## 373	0	84	64	22	66	35.8	0.545	21	0
## 374	2	105	58	40	94	34.9	0.225	25	0
## 375	2	122	52	43	158	36.2	0.816	28	0
## 376	12	140	82	43	325	39.2	0.528	58	1
## 377	0	98	82	15	84	25.2	0.299	22	0
## 378	1	87	60	37	75	37.2	0.509	22	0

## 379	4	156	75	0	0 48.3	0.238	32	1
## 380	0	93	100	39	72 43.4	1.021	35	0
## 381	1	107	72	30	82 30.8	0.821	24	0
## 382	0	105	68	22	0 20.0	0.236	22	0
## 383	1	109	60	8	182 25.4	0.947	21	0
## 384	1	90	62	18	59 25.1	1.268	25	0
## 385	1	125	70	24	110 24.3	0.221	25	0
## 386	1	119	54	13	50 22.3	0.205	24	0
## 387	5	116	74	29	0 32.3	0.660	35	1
## 388	8	105	100	36	0 43.3	0.239	45	1
## 389	5	144	82	26	285 32.0	0.452	58	1
## 390	3	100	68	23	81 31.6	0.949	28	0
## 391	1	100	66	29	196 32.0	0.444	42	0
## 392	5	166	76	0	0 45.7	0.340	27	1
## 393	1	131	64	14	415 23.7	0.389	21	0
## 394	4	116	72	12	87 22.1	0.463	37	0
## 395	4	158	78	0	0 32.9	0.803	31	1
## 396	2	127	58	24	275 27.7	1.600	25	0
## 397	3	96	56	34	115 24.7	0.944	39	0
## 398	0	131	66	40	0 34.3	0.196	22	1
## 399	3	82	70	0	0 21.1	0.389	25	0
## 400	3	193	70	31	0 34.9	0.241	25	1
## 401	4	95	64	0	0 32.0	0.161	31	1
## 402	6	137	61	0	0 24.2	0.151	55	0
## 403	5	136	84	41	88 35.0	0.286	35	1
## 404	9	72	78	25	0 31.6	0.280	38	0
## 405	5	168	64	0	0 32.9	0.135	41	1
## 406	2	123	48	32	165 42.1	0.520	26	0
## 407	4	115	72	0	0 28.9	0.376	46	1
## 408	0	101	62	0	0 21.9	0.336	25	0
## 409	8	197	74	0	0 25.9	1.191	39	1
## 410	1	172	68	49	579 42.4	0.702	28	1
## 411	6	102	90	39	0 35.7	0.674	28	0
## 412	1	112	72	30	176 34.4	0.528	25	0
## 413	1	143	84	23	310 42.4	1.076	22	0
## 414	1	143	74	22	61 26.2	0.256	21	0
## 415	0	138	60	35	167 34.6	0.534	21	1
## 416	3	173	84	33	474 35.7	0.258	22	1
## 417	1	97	68	21	0 27.2	1.095	22	0
## 418	4	144	82	32	0 38.5	0.554	37	1
## 419	1	83	68	0	0 18.2	0.624	27	0
## 420	3	129	64	29	115 26.4	0.219	28	1
## 421	1	119	88	41	170 45.3	0.507	26	0
## 422	2	94	68	18	76 26.0	0.561	21	0
## 423	0	102	64	46	78 40.6	0.496	21	0
## 424	2	115	64	22	0 30.8	0.421	21	0
## 425	8	151	78	32	210 42.9	0.516	36	1
## 426	4	184	78	39	277 37.0	0.264	31	1
## 427	0	94	0	0	0 0.0	0.256	25	0
## 428	1	181	64	30	180 34.1	0.328	38	1
## 429	0	135	94	46	145 40.6	0.284	26	0
## 430	1	95	82	25	180 35.0	0.233	43	1
## 431	2	99	0	0	0 22.2	0.108	23	0
## 432	3	89	74	16	85 30.4	0.551	38	0

## 433	1	80	74	11	60	30.0	0.527	22	0
## 434	2	139	75	0	0	25.6	0.167	29	0
## 435	1	90	68	8	0	24.5	1.138	36	0
## 436	0	141	0	0	0	42.4	0.205	29	1
## 437	12	140	85	33	0	37.4	0.244	41	0
## 438	5	147	75	0	0	29.9	0.434	28	0
## 439	1	97	70	15	0	18.2	0.147	21	0
## 440	6	107	88	0	0	36.8	0.727	31	0
## 441	0	189	104	25	0	34.3	0.435	41	1
## 442	2	83	66	23	50	32.2	0.497	22	0
## 443	4	117	64	27	120	33.2	0.230	24	0
## 444	8	108	70	0	0	30.5	0.955	33	1
## 445	4	117	62	12	0	29.7	0.380	30	1
## 446	0	180	78	63	14	59.4	2.420	25	1
## 447	1	100	72	12	70	25.3	0.658	28	0
## 448	0	95	80	45	92	36.5	0.330	26	0
## 449	0	104	64	37	64	33.6	0.510	22	1
## 450	0	120	74	18	63	30.5	0.285	26	0
## 451	1	82	64	13	95	21.2	0.415	23	0
## 452	2	134	70	0	0	28.9	0.542	23	1
## 453	0	91	68	32	210	39.9	0.381	25	0
## 454	2	119	0	0	0	19.6	0.832	72	0
## 455	2	100	54	28	105	37.8	0.498	24	0
## 456	14	175	62	30	0	33.6	0.212	38	1
## 457	1	135	54	0	0	26.7	0.687	62	0
## 458	5	86	68	28	71	30.2	0.364	24	0
## 459	10	148	84	48	237	37.6	1.001	51	1
## 460	9	134	74	33	60	25.9	0.460	81	0
## 461	9	120	72	22	56	20.8	0.733	48	0
## 462	1	71	62	0	0	21.8	0.416	26	0
## 463	8	74	70	40	49	35.3	0.705	39	0
## 464	5	88	78	30	0	27.6	0.258	37	0
## 465	10	115	98	0	0	24.0	1.022	34	0
## 466	0	124	56	13	105	21.8	0.452	21	0
## 467	0	74	52	10	36	27.8	0.269	22	0
## 468	0	97	64	36	100	36.8	0.600	25	0
## 469	8	120	0	0	0	30.0	0.183	38	1
## 470	6	154	78	41	140	46.1	0.571	27	0
## 471	1	144	82	40	0	41.3	0.607	28	0
## 472	0	137	70	38	0	33.2	0.170	22	0
## 473	0	119	66	27	0	38.8	0.259	22	0
## 474	7	136	90	0	0	29.9	0.210	50	0
## 475	4	114	64	0	0	28.9	0.126	24	0
## 476	0	137	84	27	0	27.3	0.231	59	0
## 477	2	105	80	45	191	33.7	0.711	29	1
## 478	7	114	76	17	110	23.8	0.466	31	0
## 479	8	126	74	38	75	25.9	0.162	39	0
## 480	4	132	86	31	0	28.0	0.419	63	0
## 481	3	158	70	30	328	35.5	0.344	35	1
## 482	0	123	88	37	0	35.2	0.197	29	0
## 483	4	85	58	22	49	27.8	0.306	28	0
## 484	0	84	82	31	125	38.2	0.233	23	0
## 485	0	145	0	0	0	44.2	0.630	31	1
## 486	0	135	68	42	250	42.3	0.365	24	1

## 487	1	139	62	41	480	40.7	0.536	21	0
## 488	0	173	78	32	265	46.5	1.159	58	0
## 489	4	99	72	17	0	25.6	0.294	28	0
## 490	8	194	80	0	0	26.1	0.551	67	0
## 491	2	83	65	28	66	36.8	0.629	24	0
## 492	2	89	90	30	0	33.5	0.292	42	0
## 493	4	99	68	38	0	32.8	0.145	33	0
## 494	4	125	70	18	122	28.9	1.144	45	1
## 495	3	80	0	0	0	0.0	0.174	22	0
## 496	6	166	74	0	0	26.6	0.304	66	0
## 497	5	110	68	0	0	26.0	0.292	30	0
## 498	2	81	72	15	76	30.1	0.547	25	0
## 499	7	195	70	33	145	25.1	0.163	55	1
## 500	6	154	74	32	193	29.3	0.839	39	0
## 501	2	117	90	19	71	25.2	0.313	21	0
## 502	3	84	72	32	0	37.2	0.267	28	0
## 503	6	0	68	41	0	39.0	0.727	41	1
## 504	7	94	64	25	79	33.3	0.738	41	0
## 505	3	96	78	39	0	37.3	0.238	40	0
## 506	10	75	82	0	0	33.3	0.263	38	0
## 507	0	180	90	26	90	36.5	0.314	35	1
## 508	1	130	60	23	170	28.6	0.692	21	0
## 509	2	84	50	23	76	30.4	0.968	21	0
## 510	8	120	78	0	0	25.0	0.409	64	0
## 511	12	84	72	31	0	29.7	0.297	46	1
## 512	0	139	62	17	210	22.1	0.207	21	0
## 513	9	91	68	0	0	24.2	0.200	58	0
## 514	2	91	62	0	0	27.3	0.525	22	0
## 515	3	99	54	19	86	25.6	0.154	24	0
## 516	3	163	70	18	105	31.6	0.268	28	1
## 517	9	145	88	34	165	30.3	0.771	53	1
## 518	7	125	86	0	0	37.6	0.304	51	0
## 519	13	76	60	0	0	32.8	0.180	41	0
## 520	6	129	90	7	326	19.6	0.582	60	0
## 521	2	68	70	32	66	25.0	0.187	25	0
## 522	3	124	80	33	130	33.2	0.305	26	0
## 523	6	114	0	0	0	0.0	0.189	26	0
## 524	9	130	70	0	0	34.2	0.652	45	1
## 525	3	125	58	0	0	31.6	0.151	24	0
## 526	3	87	60	18	0	21.8	0.444	21	0
## 527	1	97	64	19	82	18.2	0.299	21	0
## 528	3	116	74	15	105	26.3	0.107	24	0
## 529	0	117	66	31	188	30.8	0.493	22	0
## 530	0	111	65	0	0	24.6	0.660	31	0
## 531	2	122	60	18	106	29.8	0.717	22	0
## 532	0	107	76	0	0	45.3	0.686	24	0
## 533	1	86	66	52	65	41.3	0.917	29	0
## 534	6	91	0	0	0	29.8	0.501	31	0
## 535	1	77	56	30	56	33.3	1.251	24	0
## 536	4	132	0	0	0	32.9	0.302	23	1
## 537	0	105	90	0	0	29.6	0.197	46	0
## 538	0	57	60	0	0	21.7	0.735	67	0
## 539	0	127	80	37	210	36.3	0.804	23	0
## 540	3	129	92	49	155	36.4	0.968	32	1

## 541	8	100	74	40	215	39.4	0.661	43	1
## 542	3	128	72	25	190	32.4	0.549	27	1
## 543	10	90	85	32	0	34.9	0.825	56	1
## 544	4	84	90	23	56	39.5	0.159	25	0
## 545	1	88	78	29	76	32.0	0.365	29	0
## 546	8	186	90	35	225	34.5	0.423	37	1
## 547	5	187	76	27	207	43.6	1.034	53	1
## 548	4	131	68	21	166	33.1	0.160	28	0
## 549	1	164	82	43	67	32.8	0.341	50	0
## 550	4	189	110	31	0	28.5	0.680	37	0
## 551	1	116	70	28	0	27.4	0.204	21	0
## 552	3	84	68	30	106	31.9	0.591	25	0
## 553	6	114	88	0	0	27.8	0.247	66	0
## 554	1	88	62	24	44	29.9	0.422	23	0
## 555	1	84	64	23	115	36.9	0.471	28	0
## 556	7	124	70	33	215	25.5	0.161	37	0
## 557	1	97	70	40	0	38.1	0.218	30	0
## 558	8	110	76	0	0	27.8	0.237	58	0
## 559	11	103	68	40	0	46.2	0.126	42	0
## 560	11	85	74	0	0	30.1	0.300	35	0
## 561	6	125	76	0	0	33.8	0.121	54	1
## 562	0	198	66	32	274	41.3	0.502	28	1
## 563	1	87	68	34	77	37.6	0.401	24	0
## 564	6	99	60	19	54	26.9	0.497	32	0
## 565	0	91	80	0	0	32.4	0.601	27	0
## 566	2	95	54	14	88	26.1	0.748	22	0
## 567	1	99	72	30	18	38.6	0.412	21	0
## 568	6	92	62	32	126	32.0	0.085	46	0
## 569	4	154	72	29	126	31.3	0.338	37	0
## 570	0	121	66	30	165	34.3	0.203	33	1
## 571	3	78	70	0	0	32.5	0.270	39	0
## 572	2	130	96	0	0	22.6	0.268	21	0
## 573	3	111	58	31	44	29.5	0.430	22	0
## 574	2	98	60	17	120	34.7	0.198	22	0
## 575	1	143	86	30	330	30.1	0.892	23	0
## 576	1	119	44	47	63	35.5	0.280	25	0
## 577	6	108	44	20	130	24.0	0.813	35	0
## 578	2	118	80	0	0	42.9	0.693	21	1
## 579	10	133	68	0	0	27.0	0.245	36	0
## 580	2	197	70	99	0	34.7	0.575	62	1
## 581	0	151	90	46	0	42.1	0.371	21	1
## 582	6	109	60	27	0	25.0	0.206	27	0
## 583	12	121	78	17	0	26.5	0.259	62	0
## 584	8	100	76	0	0	38.7	0.190	42	0
## 585	8	124	76	24	600	28.7	0.687	52	1
## 586	1	93	56	11	0	22.5	0.417	22	0
## 587	8	143	66	0	0	34.9	0.129	41	1
## 588	6	103	66	0	0	24.3	0.249	29	0
## 589	3	176	86	27	156	33.3	1.154	52	1
## 590	0	73	0	0	0	21.1	0.342	25	0
## 591	11	111	84	40	0	46.8	0.925	45	1
## 592	2	112	78	50	140	39.4	0.175	24	0
## 593	3	132	80	0	0	34.4	0.402	44	1
## 594	2	82	52	22	115	28.5	1.699	25	0

## 595	6	123	72	45	230	33.6	0.733	34	0
## 596	0	188	82	14	185	32.0	0.682	22	1
## 597	0	67	76	0	0	45.3	0.194	46	0
## 598	1	89	24	19	25	27.8	0.559	21	0
## 599	1	173	74	0	0	36.8	0.088	38	1
## 600	1	109	38	18	120	23.1	0.407	26	0
## 601	1	108	88	19	0	27.1	0.400	24	0
## 602	6	96	0	0	0	23.7	0.190	28	0
## 603	1	124	74	36	0	27.8	0.100	30	0
## 604	7	150	78	29	126	35.2	0.692	54	1
## 605	4	183	0	0	0	28.4	0.212	36	1
## 606	1	124	60	32	0	35.8	0.514	21	0
## 607	1	181	78	42	293	40.0	1.258	22	1
## 608	1	92	62	25	41	19.5	0.482	25	0
## 609	0	152	82	39	272	41.5	0.270	27	0
## 610	1	111	62	13	182	24.0	0.138	23	0
## 611	3	106	54	21	158	30.9	0.292	24	0
## 612	3	174	58	22	194	32.9	0.593	36	1
## 613	7	168	88	42	321	38.2	0.787	40	1
## 614	6	105	80	28	0	32.5	0.878	26	0
## 615	11	138	74	26	144	36.1	0.557	50	1
## 616	3	106	72	0	0	25.8	0.207	27	0
## 617	6	117	96	0	0	28.7	0.157	30	0
## 618	2	68	62	13	15	20.1	0.257	23	0
## 619	9	112	82	24	0	28.2	1.282	50	1
## 620	0	119	0	0	0	32.4	0.141	24	1
## 621	2	112	86	42	160	38.4	0.246	28	0
## 622	2	92	76	20	0	24.2	1.698	28	0
## 623	6	183	94	0	0	40.8	1.461	45	0
## 624	0	94	70	27	115	43.5	0.347	21	0
## 625	2	108	64	0	0	30.8	0.158	21	0
## 626	4	90	88	47	54	37.7	0.362	29	0
## 627	0	125	68	0	0	24.7	0.206	21	0
## 628	0	132	78	0	0	32.4	0.393	21	0
## 629	5	128	80	0	0	34.6	0.144	45	0
## 630	4	94	65	22	0	24.7	0.148	21	0
## 631	7	114	64	0	0	27.4	0.732	34	1
## 632	0	102	78	40	90	34.5	0.238	24	0
## 633	2	111	60	0	0	26.2	0.343	23	0
## 634	1	128	82	17	183	27.5	0.115	22	0
## 635	10	92	62	0	0	25.9	0.167	31	0
## 636	13	104	72	0	0	31.2	0.465	38	1
## 637	5	104	74	0	0	28.8	0.153	48	0
## 638	2	94	76	18	66	31.6	0.649	23	0
## 639	7	97	76	32	91	40.9	0.871	32	1
## 640	1	100	74	12	46	19.5	0.149	28	0
## 641	0	102	86	17	105	29.3	0.695	27	0
## 642	4	128	70	0	0	34.3	0.303	24	0
## 643	6	147	80	0	0	29.5	0.178	50	1
## 644	4	90	0	0	0	28.0	0.610	31	0
## 645	3	103	72	30	152	27.6	0.730	27	0
## 646	2	157	74	35	440	39.4	0.134	30	0
## 647	1	167	74	17	144	23.4	0.447	33	1
## 648	0	179	50	36	159	37.8	0.455	22	1

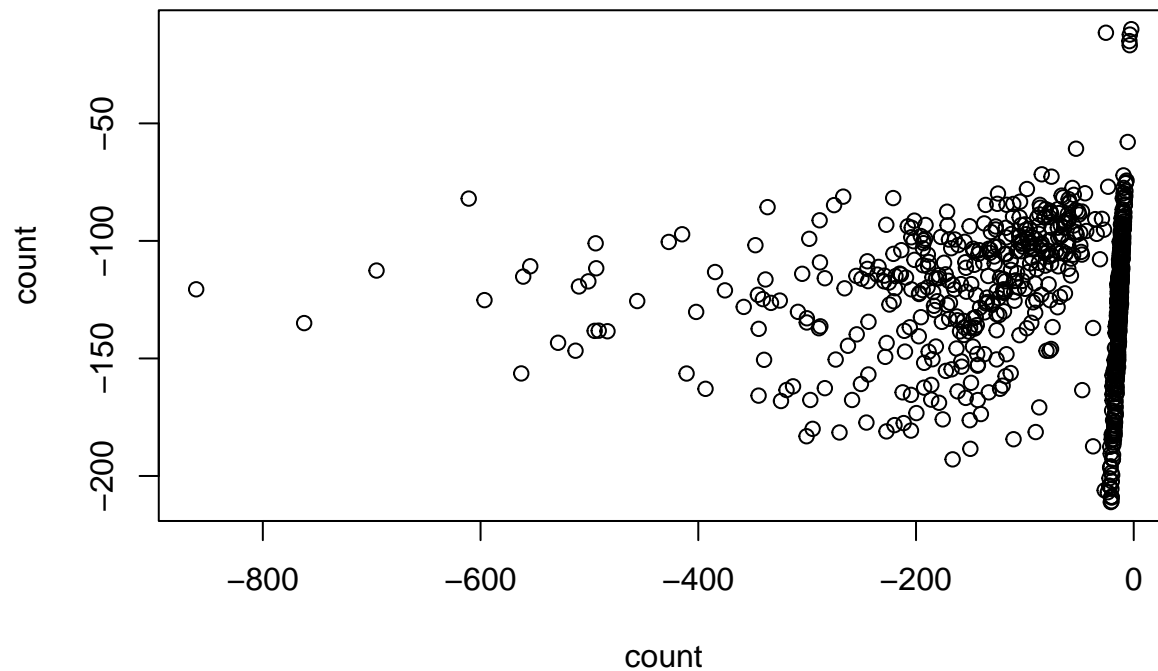
## 649	11	136	84	35	130	28.3	0.260	42	1
## 650	0	107	60	25	0	26.4	0.133	23	0
## 651	1	91	54	25	100	25.2	0.234	23	0
## 652	1	117	60	23	106	33.8	0.466	27	0
## 653	5	123	74	40	77	34.1	0.269	28	0
## 654	2	120	54	0	0	26.8	0.455	27	0
## 655	1	106	70	28	135	34.2	0.142	22	0
## 656	2	155	52	27	540	38.7	0.240	25	1
## 657	2	101	58	35	90	21.8	0.155	22	0
## 658	1	120	80	48	200	38.9	1.162	41	0
## 659	11	127	106	0	0	39.0	0.190	51	0
## 660	3	80	82	31	70	34.2	1.292	27	1
## 661	10	162	84	0	0	27.7	0.182	54	0
## 662	1	199	76	43	0	42.9	1.394	22	1
## 663	8	167	106	46	231	37.6	0.165	43	1
## 664	9	145	80	46	130	37.9	0.637	40	1
## 665	6	115	60	39	0	33.7	0.245	40	1
## 666	1	112	80	45	132	34.8	0.217	24	0
## 667	4	145	82	18	0	32.5	0.235	70	1
## 668	10	111	70	27	0	27.5	0.141	40	1
## 669	6	98	58	33	190	34.0	0.430	43	0
## 670	9	154	78	30	100	30.9	0.164	45	0
## 671	6	165	68	26	168	33.6	0.631	49	0
## 672	1	99	58	10	0	25.4	0.551	21	0
## 673	10	68	106	23	49	35.5	0.285	47	0
## 674	3	123	100	35	240	57.3	0.880	22	0
## 675	8	91	82	0	0	35.6	0.587	68	0
## 676	6	195	70	0	0	30.9	0.328	31	1
## 677	9	156	86	0	0	24.8	0.230	53	1
## 678	0	93	60	0	0	35.3	0.263	25	0
## 679	3	121	52	0	0	36.0	0.127	25	1
## 680	2	101	58	17	265	24.2	0.614	23	0
## 681	2	56	56	28	45	24.2	0.332	22	0
## 682	0	162	76	36	0	49.6	0.364	26	1
## 683	0	95	64	39	105	44.6	0.366	22	0
## 684	4	125	80	0	0	32.3	0.536	27	1
## 685	5	136	82	0	0	0.0	0.640	69	0
## 686	2	129	74	26	205	33.2	0.591	25	0
## 687	3	130	64	0	0	23.1	0.314	22	0
## 688	1	107	50	19	0	28.3	0.181	29	0
## 689	1	140	74	26	180	24.1	0.828	23	0
## 690	1	144	82	46	180	46.1	0.335	46	1
## 691	8	107	80	0	0	24.6	0.856	34	0
## 692	13	158	114	0	0	42.3	0.257	44	1
## 693	2	121	70	32	95	39.1	0.886	23	0
## 694	7	129	68	49	125	38.5	0.439	43	1
## 695	2	90	60	0	0	23.5	0.191	25	0
## 696	7	142	90	24	480	30.4	0.128	43	1
## 697	3	169	74	19	125	29.9	0.268	31	1
## 698	0	99	0	0	0	25.0	0.253	22	0
## 699	4	127	88	11	155	34.5	0.598	28	0
## 700	4	118	70	0	0	44.5	0.904	26	0
## 701	2	122	76	27	200	35.9	0.483	26	0
## 702	6	125	78	31	0	27.6	0.565	49	1

## 703	1	168	88	29	0	35.0	0.905	52	1
## 704	2	129	0	0	0	38.5	0.304	41	0
## 705	4	110	76	20	100	28.4	0.118	27	0
## 706	6	80	80	36	0	39.8	0.177	28	0
## 707	10	115	0	0	0	0.0	0.261	30	1
## 708	2	127	46	21	335	34.4	0.176	22	0
## 709	9	164	78	0	0	32.8	0.148	45	1
## 710	2	93	64	32	160	38.0	0.674	23	1
## 711	3	158	64	13	387	31.2	0.295	24	0
## 712	5	126	78	27	22	29.6	0.439	40	0
## 713	10	129	62	36	0	41.2	0.441	38	1
## 714	0	134	58	20	291	26.4	0.352	21	0
## 715	3	102	74	0	0	29.5	0.121	32	0
## 716	7	187	50	33	392	33.9	0.826	34	1
## 717	3	173	78	39	185	33.8	0.970	31	1
## 718	10	94	72	18	0	23.1	0.595	56	0
## 719	1	108	60	46	178	35.5	0.415	24	0
## 720	5	97	76	27	0	35.6	0.378	52	1
## 721	4	83	86	19	0	29.3	0.317	34	0
## 722	1	114	66	36	200	38.1	0.289	21	0
## 723	1	149	68	29	127	29.3	0.349	42	1
## 724	5	117	86	30	105	39.1	0.251	42	0
## 725	1	111	94	0	0	32.8	0.265	45	0
## 726	4	112	78	40	0	39.4	0.236	38	0
## 727	1	116	78	29	180	36.1	0.496	25	0
## 728	0	141	84	26	0	32.4	0.433	22	0
## 729	2	175	88	0	0	22.9	0.326	22	0
## 730	2	92	52	0	0	30.1	0.141	22	0
## 731	3	130	78	23	79	28.4	0.323	34	1
## 732	8	120	86	0	0	28.4	0.259	22	1
## 733	2	174	88	37	120	44.5	0.646	24	1
## 734	2	106	56	27	165	29.0	0.426	22	0
## 735	2	105	75	0	0	23.3	0.560	53	0
## 736	4	95	60	32	0	35.4	0.284	28	0
## 737	0	126	86	27	120	27.4	0.515	21	0
## 738	8	65	72	23	0	32.0	0.600	42	0
## 739	2	99	60	17	160	36.6	0.453	21	0
## 740	1	102	74	0	0	39.5	0.293	42	1
## 741	11	120	80	37	150	42.3	0.785	48	1
## 742	3	102	44	20	94	30.8	0.400	26	0
## 743	1	109	58	18	116	28.5	0.219	22	0
## 744	9	140	94	0	0	32.7	0.734	45	1
## 745	13	153	88	37	140	40.6	1.174	39	0
## 746	12	100	84	33	105	30.0	0.488	46	0
## 747	1	147	94	41	0	49.3	0.358	27	1
## 748	1	81	74	41	57	46.3	1.096	32	0
## 749	3	187	70	22	200	36.4	0.408	36	1
## 750	6	162	62	0	0	24.3	0.178	50	1
## 751	4	136	70	0	0	31.2	1.182	22	1
## 752	1	121	78	39	74	39.0	0.261	28	0
## 753	3	108	62	24	0	26.0	0.223	25	0
## 754	0	181	88	44	510	43.3	0.222	26	1
## 755	8	154	78	32	0	32.4	0.443	45	1
## 756	1	128	88	39	110	36.5	1.057	37	1

```
## 757      7      137      90      41      0 32.0      0.391 39      0
## 758      0      123      72       0      0 36.3      0.258 52      1
## 759      1      106      76       0      0 37.5      0.197 26      0
## 760      6      190      92       0      0 35.5      0.278 66      1
## 761      2       88      58      26     16 28.4      0.766 22      0
## 762      9      170      74      31      0 44.0      0.403 43      1
## 763      9       89      62       0      0 22.5      0.142 33      0
## 764     10      101      76      48     180 32.9      0.171 63      0
## 765      2      122      70      27      0 36.8      0.340 27      0
## 766      5      121      72      23     112 26.2      0.245 30      0
## 767      1      126      60       0      0 30.1      0.349 47      1
## 768      1       93      70      31      0 30.4      0.315 23      0
```

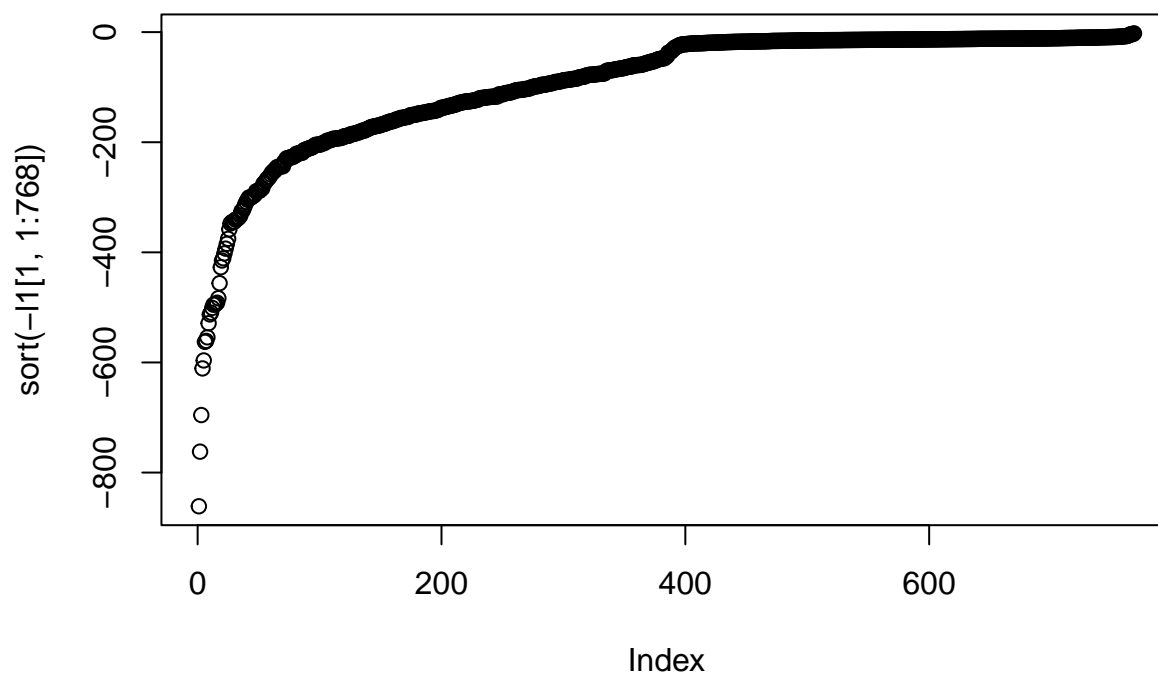
```
#get the numeric values
a <- PimaIndiansDiabetes[, sapply(PimaIndiansDiabetes, is.numeric)]
ev = eigen(cov(a))$values
eve = eigen(cov(a))$vectors
l1 = eve[1:9, 1] %*% t(a)
l2 = eve[1:9, 2] %*% t(a)
plot(-l1[1, 1:768], l2[1, 1:768], main = "PC1 vs. PC2 Correlation", xlab = "count", ylab = "count")
```

PC1 vs. PC2 Correlation



```
plot(sort(-l1[1, 1:768]), main = "PC1 Plot")
```

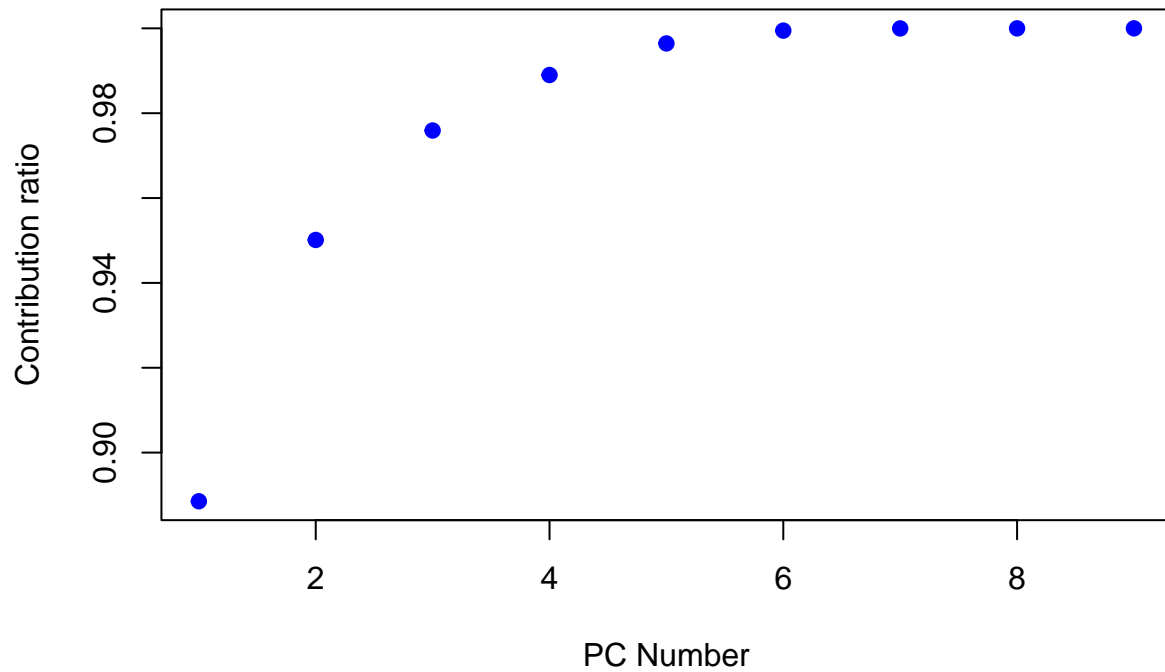
PC1 Plot



```
## contribution ratio
Cr = rep(0, length(ev))
for (k in 1:length(ev))
{
  Cr[k] = sum(ev[1:k]) / sum(ev)
}

plot(Cr, main = "Contribution Ratio Plot", xlab = "PC Number", ylab = "Contribution ratio", pch = 19, col = "red",
      abline(h = 0.8, col = "red"))
```

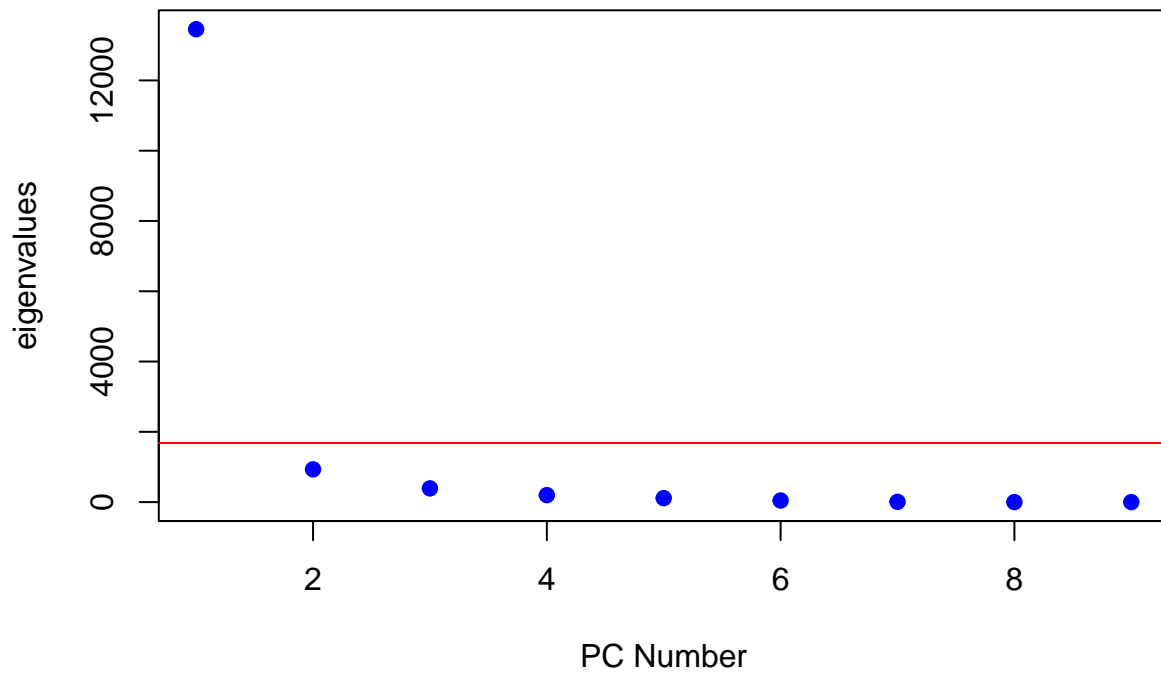
Contribution Ratio Plot



```
## overall threshold
```

```
plot(ev, main = "Overall Threshold Plot", xlab = "PC Number", ylab = "eigenvalues", pch = 19, col = "blue",  
abline(h = mean(ev), col = "red"))
```

Overall Threshold Plot



```

#here we find the numeric values of PC1 and PC2
load <- eve[, 1]
load2 <- eve[, 2]
#here we are finding the column names so we can link them to the components
individualFeatures <- colnames(a)
#here we are looking to see what features these numbers correspond
PC1 <- data.frame(Feature = individualFeatures, Loading = load)
PC2 <- data.frame(Feature = individualFeatures, Loading = load2)

#print
print(PC1)

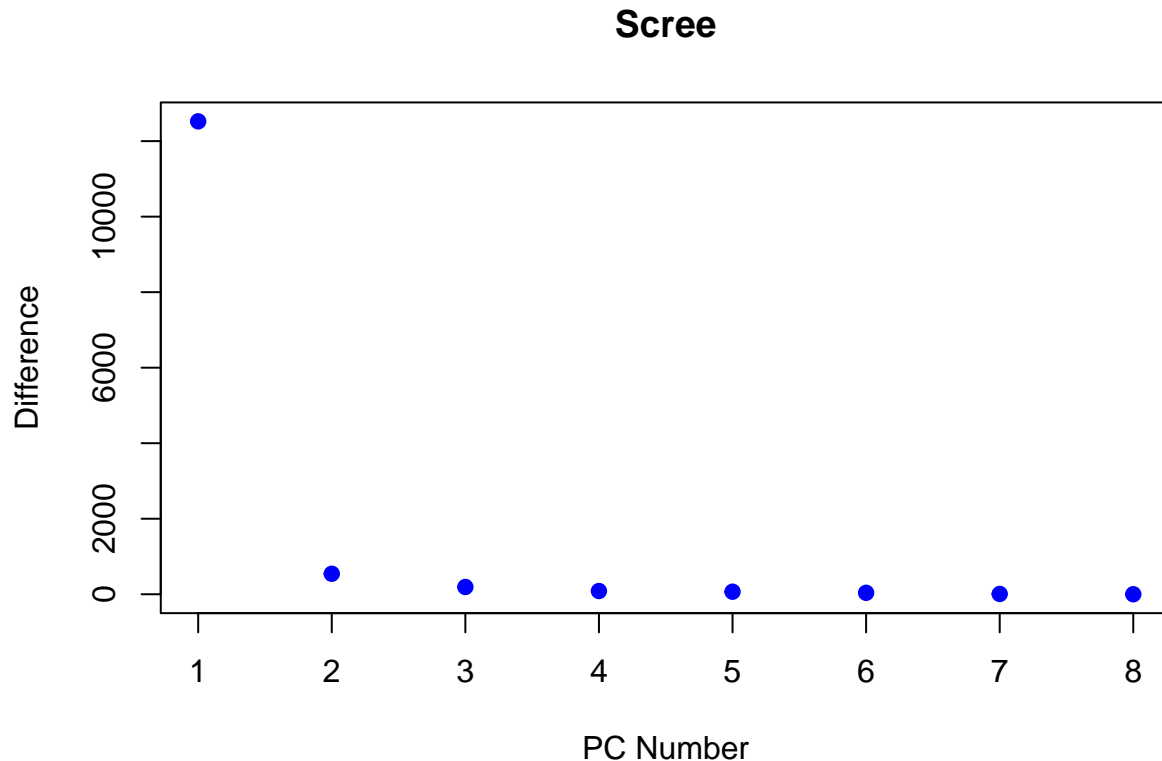
##      Feature      Loading
## 1 pregnant -0.0020217488
## 2  glucose  0.0978118564
## 3 pressure  0.0160930708
## 4  triceps  0.0607566786
## 5  insulin  0.9931106435
## 6    mass   0.0140108503
## 7 pedigree  0.0005371689
## 8    age   -0.0035646812
## 9 diabetes  0.0005853255

print(PC2)

##      Feature      Loading
## 1 pregnant -0.0226500774
## 2  glucose -0.9721857784
## 3 pressure -0.1419012985
## 4  triceps  0.0578559304
## 5  insulin  0.0946290072
## 6    mass  -0.0469772538
## 7 pedigree -0.0008169207
## 8    age   -0.1401683828
## 9 diabetes -0.0070103371

## Scree
plot(ev[1:8]-ev[2:9], main = "Scree", xlab = "PC Number", ylab="Difference", pch = 19,col="blue")

```



PCA Plots (PC1 vs. PC2, PC1): Plot 1 shows the correlation between PC1 and PC2 in this case we know that PC1 is mainly driven by insulin levels and PC2 is driven by glucose. The cluster at the origin suggests that most people have average levels of both factors. Those straying right on the x-axis indicate higher insulin, and those straying up on the y-axis indicate higher glucose levels. We can see from this plot the individuals that are at risk for both factors are the most likely to be at risk.

The second plot shows us the importance of the insulin factor as insulin is by far the highest factor in PC1. Based on the plot we can see that insulin and PC1 values correlate quite directly, proving the importance of the insulin factor.

Contribution Ratio: In the contribution ratio plot shows us how many principal components are needed to get a majority of the variance. We can see a fast incline from the first 2 points and then the rest of the components don't provide much of an increase. This shows us that PC1 and PC2 are the main factors for the variance. Overall from this plot we can see that after the first 6 components the plot flattens out.

Overall Threshold: For the overall threshold we look at the eigenvalues and as shown here there is a large drop from PC1 to PC2. The line shows the mean and those above it have significance to the variance which as we can see only PC1 is above the line, and in far second place PC2 is closest to the line out of the remaining components.

Scree Plot: The scree plot shows us a very similar thing as the overall threshold plot. Instead of looking at the big drop in eigenvalues to interpret the amount of components needed, the Scree plot gives us a bigger overview of the eigenvalues for our dataset. This shows us the rate of variation as you go from one component to the next. As we can see here the output is generally the same showing the eigenvalue for PC1 to be by far the highest.