Intro to R

#### Execute the following cells to load the libraries

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#### Load the food texture dataset

file = 'Codes/Data/food-texture.csv'  
#file = 'Data/food-texture.csv'  
foodData = read.csv(file, header = TRUE, row.names = 1, stringsAsFactors = FALSE)  
head(foodData, 2)

## Oil Density Crispy Fracture Hardness  
## B110 16.5 2955 10 23 97  
## B136 17.7 2660 14 9 139

str(foodData)

## 'data.frame': 50 obs. of 5 variables:  
## $ Oil : num 16.5 17.7 16.2 16.7 16.3 19.1 18.4 17.5 15.7 16.4 ...  
## $ Density : int 2955 2660 2870 2920 2975 2790 2750 2770 2955 2945 ...  
## $ Crispy : int 10 14 12 10 11 13 13 10 11 11 ...  
## $ Fracture: int 23 9 17 31 26 16 17 26 23 24 ...  
## $ Hardness: int 97 139 143 95 143 189 114 63 123 132 ...

#### Modify Crispy column to reflect high (0) or low (1) crispiness

foodData = foodData %>% mutate(Crispylevel = ifelse(Crispy > 11, 'High', 'Low'))

##### Change Crispy and Crispylevel columns to factor (categorical) type

# Continuous features -> Oil, Density, Hardness, Fracture  
# Categorical features -> Crispy (8 levels 8 through 15) and Crispylevel (2 levels 0 and 1)  
# Crispy is a categorical feature with an order  
str(foodData)

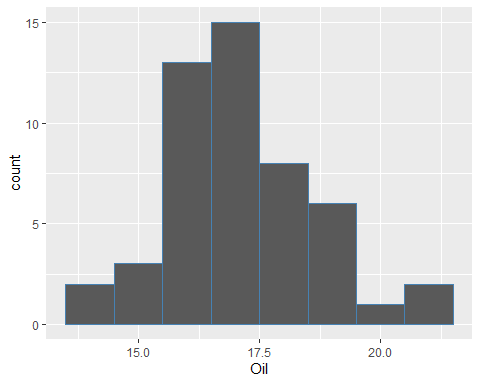
## 'data.frame': 50 obs. of 6 variables:  
## $ Oil : num 16.5 17.7 16.2 16.7 16.3 19.1 18.4 17.5 15.7 16.4 ...  
## $ Density : int 2955 2660 2870 2920 2975 2790 2750 2770 2955 2945 ...  
## $ Crispy : int 10 14 12 10 11 13 13 10 11 11 ...  
## $ Fracture : int 23 9 17 31 26 16 17 26 23 24 ...  
## $ Hardness : int 97 139 143 95 143 189 114 63 123 132 ...  
## $ Crispylevel: chr "Low" "High" "High" "Low" ...

categorical\_cols = c('Crispy', 'Crispylevel')  
foodData[categorical\_cols] = lapply(foodData[categorical\_cols], as.factor)  
str(foodData)

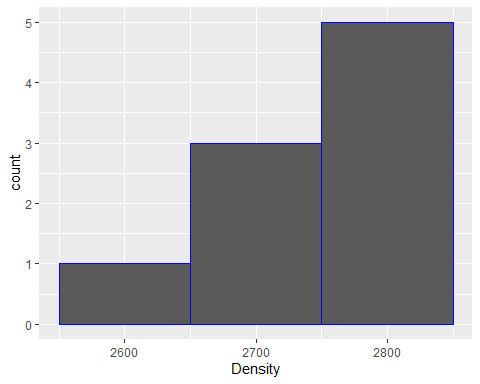
## 'data.frame': 50 obs. of 6 variables:  
## $ Oil : num 16.5 17.7 16.2 16.7 16.3 19.1 18.4 17.5 15.7 16.4 ...  
## $ Density : int 2955 2660 2870 2920 2975 2790 2750 2770 2955 2945 ...  
## $ Crispy : Factor w/ 9 levels "7","8","9","10",..: 4 8 6 4 5 7 7 4 5 5 ...  
## $ Fracture : int 23 9 17 31 26 16 17 26 23 24 ...  
## $ Hardness : int 97 139 143 95 143 189 114 63 123 132 ...  
## $ Crispylevel: Factor w/ 2 levels "High","Low": 2 1 1 2 2 1 1 2 2 2 ...

#### Visualize the Oil Percentage feature (continuous) using a histogram

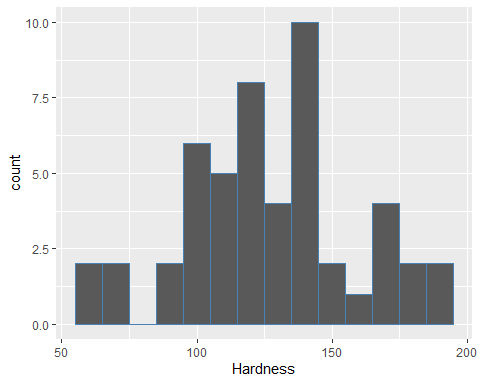
p = ggplot(data = foodData) +  
 geom\_histogram(aes(x = Oil), binwidth = 1, color = 'steelblue')  
p



p = ggplot(data = foodData[foodData$Crispy == 13, ]) +  
 geom\_histogram(aes(x = Density), binwidth = 100, color = 'blue')  
p

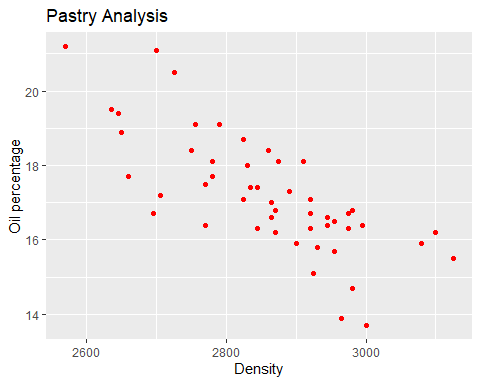


p = ggplot(data = foodData) +  
 geom\_histogram(aes(x = Hardness), binwidth = 10, color = 'steelblue')  
p



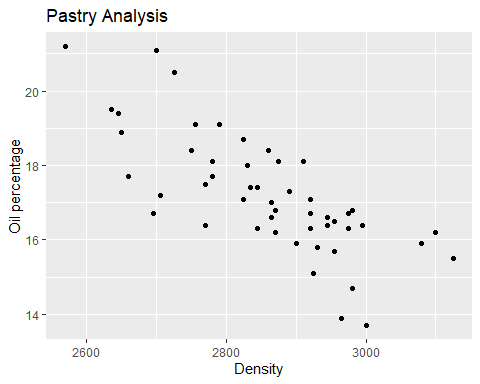
#### Scatter plot between Oil Percentage and Density (relating two continuous variables)

p = ggplot(data = foodData) +  
 geom\_point(aes(x = Density, y = Oil), color = 'red') +  
 labs(x = 'Density', y = 'Oil percentage', title = 'Pastry Analysis')  
p

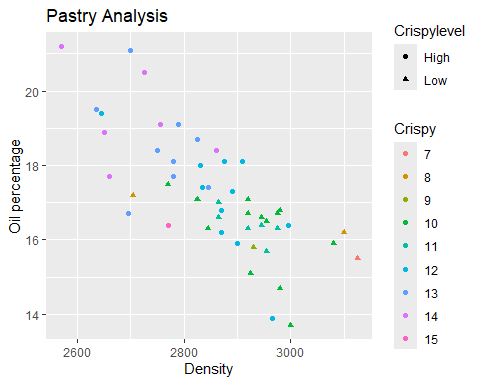


#### Scatter plot between OilPercentage and Density color coded with Crispylevel

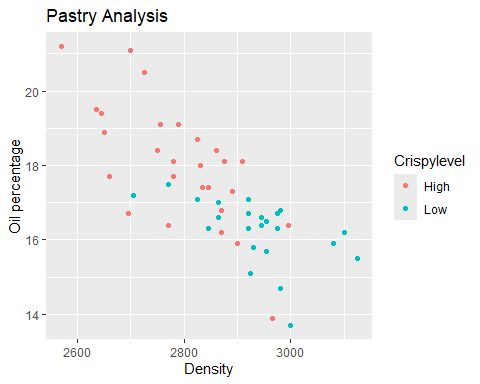
p = ggplot(data = foodData) +  
 geom\_point(aes(x = Density, y = Oil)) +  
 labs(x = 'Density', y = 'Oil percentage', title = 'Pastry Analysis')  
p



p = ggplot(data = foodData) +  
 geom\_point(aes(x = Density, y = Oil, color = Crispy, shape= Crispylevel)) +  
 labs(x = 'Density', y = 'Oil percentage', title = 'Pastry Analysis')  
p



p = ggplot(data = foodData) +  
 geom\_point(aes(x = Density, y = Oil, color = Crispylevel)) +  
 labs(x = 'Density', y = 'Oil percentage', title = 'Pastry Analysis')  
p



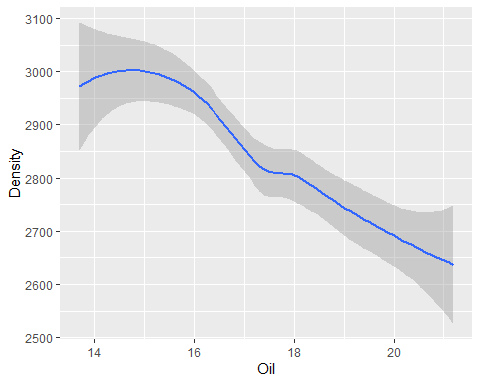
#### Scatter plot between Density and Hardness

#### Scatter plot between Density and Fracture

#### Smooth line plot using ggplot

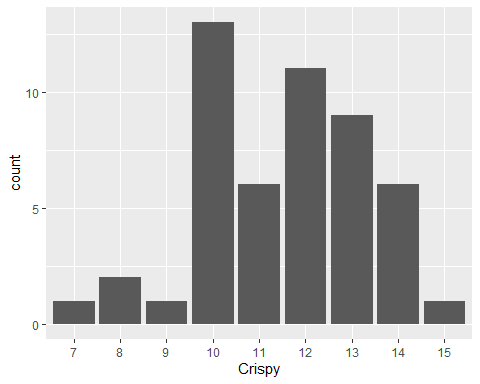
p = ggplot(data = foodData) +  
 geom\_smooth(aes(x = Oil, y = Density))  
p

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'



#### Barplot for Crispy (categorical variable)

p = ggplot(data = foodData) +  
 geom\_bar(aes(x = Crispy))  
p

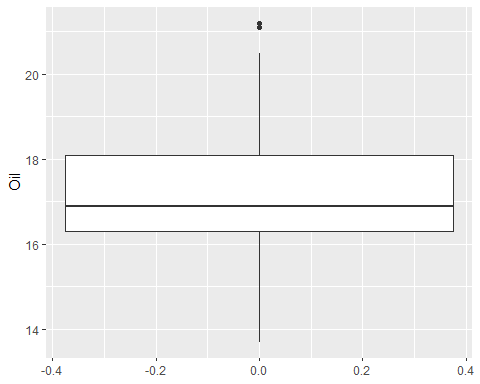


#### In-built functions for dataframes

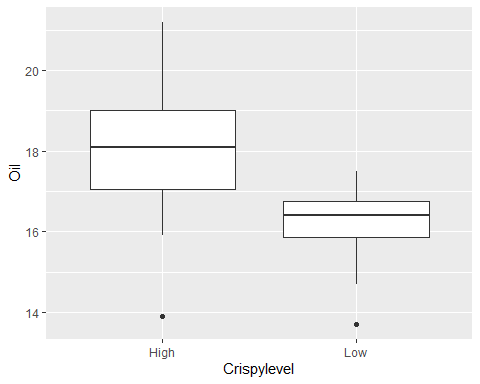
# Mean oil percentage across all samples  
  
# Mean-centering of OilPercentage  
  
# Sum of the squared deviation from the mean  
  
# Average of the squared deviation from the mean  
  
# Variance of OilPercentage  
  
# Standard deviation of OilPercentage

#### Box plot using ggplot with color coding

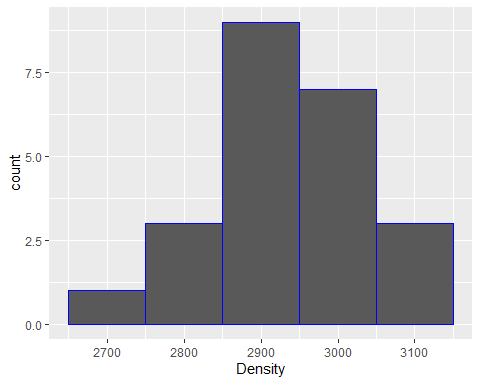
p = ggplot(data = foodData) +  
 geom\_boxplot(aes(y = Oil))  
p



p = ggplot(data = foodData) +  
 geom\_boxplot(aes(x = Crispylevel, y = Oil))  
p



p = ggplot(data = foodData[foodData$Crispylevel == 'Low', ]) +  
 geom\_histogram(aes(x = Density), binwidth = 100, color = 'blue')  
p



#####percentile and quantiles

# what is a percentile? sometimes also reffered to

# as the quantile

foodData %>% select(Oil)

## Oil  
## B110 16.5  
## B136 17.7  
## B171 16.2  
## B192 16.7  
## B225 16.3  
## B237 19.1  
## B261 18.4  
## B264 17.5  
## B353 15.7  
## B360 16.4  
## B366 18.0  
## B377 17.4  
## B391 18.4  
## B397 13.9  
## B404 15.8  
## B437 16.4  
## B445 18.9  
## B462 17.3  
## B485 16.7  
## B488 19.1  
## B502 13.7  
## B554 14.7  
## B556 18.1  
## B575 17.2  
## B576 18.7  
## B605 18.1  
## B612 16.6  
## B615 17.1  
## B649 17.4  
## B665 19.4  
## B674 15.9  
## B692 17.1  
## B694 15.5  
## B719 17.7  
## B727 15.9  
## B758 21.2  
## B776 19.5  
## B799 20.5  
## B836 17.0  
## B848 16.7  
## B861 16.8  
## B869 16.8  
## B876 16.3  
## B882 16.2  
## B889 18.1  
## B907 16.6  
## B911 16.4  
## B923 15.1  
## B971 21.1  
## B998 16.3

median(foodData$Oil) # 50 th percentile or 0.5th quantile

## [1] 16.9

mean(foodData$Oil)

## [1] 17.202

foodData %>% select(Oil)

## Oil  
## B110 16.5  
## B136 17.7  
## B171 16.2  
## B192 16.7  
## B225 16.3  
## B237 19.1  
## B261 18.4  
## B264 17.5  
## B353 15.7  
## B360 16.4  
## B366 18.0  
## B377 17.4  
## B391 18.4  
## B397 13.9  
## B404 15.8  
## B437 16.4  
## B445 18.9  
## B462 17.3  
## B485 16.7  
## B488 19.1  
## B502 13.7  
## B554 14.7  
## B556 18.1  
## B575 17.2  
## B576 18.7  
## B605 18.1  
## B612 16.6  
## B615 17.1  
## B649 17.4  
## B665 19.4  
## B674 15.9  
## B692 17.1  
## B694 15.5  
## B719 17.7  
## B727 15.9  
## B758 21.2  
## B776 19.5  
## B799 20.5  
## B836 17.0  
## B848 16.7  
## B861 16.8  
## B869 16.8  
## B876 16.3  
## B882 16.2  
## B889 18.1  
## B907 16.6  
## B911 16.4  
## B923 15.1  
## B971 21.1  
## B998 16.3

Oil\_median=median(foodData$Oil) # 50 th percentile or 0.5th quantile  
#mean(foodData$Oil)  
print(Oil\_median)

## [1] 16.9

# find the fraction of samples whose oil percentage is less than or equal to median oil percentage (50th percentile)=0.5

#P(Random sample’s oil percentage <=median oil percentage)= 0.5

foodData %>% select(Oil)

## Oil  
## B110 16.5  
## B136 17.7  
## B171 16.2  
## B192 16.7  
## B225 16.3  
## B237 19.1  
## B261 18.4  
## B264 17.5  
## B353 15.7  
## B360 16.4  
## B366 18.0  
## B377 17.4  
## B391 18.4  
## B397 13.9  
## B404 15.8  
## B437 16.4  
## B445 18.9  
## B462 17.3  
## B485 16.7  
## B488 19.1  
## B502 13.7  
## B554 14.7  
## B556 18.1  
## B575 17.2  
## B576 18.7  
## B605 18.1  
## B612 16.6  
## B615 17.1  
## B649 17.4  
## B665 19.4  
## B674 15.9  
## B692 17.1  
## B694 15.5  
## B719 17.7  
## B727 15.9  
## B758 21.2  
## B776 19.5  
## B799 20.5  
## B836 17.0  
## B848 16.7  
## B861 16.8  
## B869 16.8  
## B876 16.3  
## B882 16.2  
## B889 18.1  
## B907 16.6  
## B911 16.4  
## B923 15.1  
## B971 21.1  
## B998 16.3

Oil\_median=median(foodData$Oil) # 50 th percentile or 0.5th quantile  
#mean(foodData$Oil)  
print(Oil\_median)

## [1] 16.9

#foodData %>% filter(Oil <= Oil\_median) %>% select(Oil)  
nrow(foodData %>% filter(Oil <= Oil\_median) %>% select(Oil))/nrow(foodData)

## [1] 0.5

#foodData$Oil --> vector output  
#foodData$Oil <=Oil\_median --> true / false o/p  
mean(foodData$Oil <=Oil\_median)

## [1] 0.5

#Diabetes

#### Load the food Diabetes dataset

file = 'Codes/Data/diabetes.csv'  
#file = 'Data/food-texture.csv'  
diabetData = read.csv(file, header = TRUE, stringsAsFactors = FALSE)  
head(diabetData, 2)

## Pregnancies Glucose BloodPressure SkinThickness Insulin BMI  
## 1 6 148 72 35 0 33.6  
## 2 1 85 66 29 0 26.6  
## DiabetesPedigreeFunction Age Outcome  
## 1 0.627 50 1  
## 2 0.351 31 0

str(diabetData)

## 'data.frame': 768 obs. of 9 variables:  
## $ Pregnancies : int 6 1 8 1 0 5 3 10 2 8 ...  
## $ Glucose : int 148 85 183 89 137 116 78 115 197 125 ...  
## $ BloodPressure : int 72 66 64 66 40 74 50 0 70 96 ...  
## $ SkinThickness : int 35 29 0 23 35 0 32 0 45 0 ...  
## $ Insulin : int 0 0 0 94 168 0 88 0 543 0 ...  
## $ BMI : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...  
## $ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...  
## $ Age : int 50 31 32 21 33 30 26 29 53 54 ...  
## $ Outcome : int 1 0 1 0 1 0 1 0 1 1 ...

categorical\_cols=c('Outcome')  
foodData[categorical\_cols] = lapply(diabetData[categorical\_cols], as.factor)

## Warning in `[<-.data.frame`(`\*tmp\*`, categorical\_cols, value = list(Outcome =  
## structure(c(2L, : replacement element 1 has 768 rows to replace 50 rows

str(foodData)

## 'data.frame': 50 obs. of 7 variables:  
## $ Oil : num 16.5 17.7 16.2 16.7 16.3 19.1 18.4 17.5 15.7 16.4 ...  
## $ Density : int 2955 2660 2870 2920 2975 2790 2750 2770 2955 2945 ...  
## $ Crispy : Factor w/ 9 levels "7","8","9","10",..: 4 8 6 4 5 7 7 4 5 5 ...  
## $ Fracture : int 23 9 17 31 26 16 17 26 23 24 ...  
## $ Hardness : int 97 139 143 95 143 189 114 63 123 132 ...  
## $ Crispylevel: Factor w/ 2 levels "High","Low": 2 1 1 2 2 1 1 2 2 2 ...  
## $ Outcome : Factor w/ 2 levels "0","1": 2 1 2 1 2 1 2 1 2 2 ...

#diabetData = diabetData %>% mutate(Plevel = ifelse(Pregnancies > 10, 'High', 'Low'))

str(diabetData)

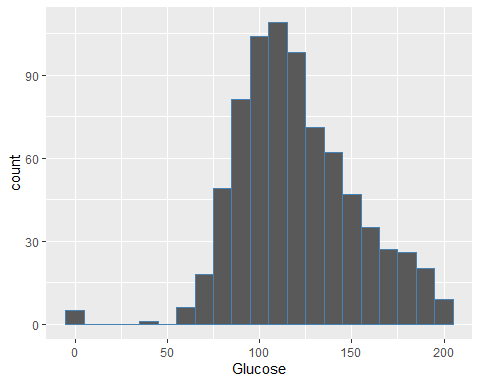
## 'data.frame': 768 obs. of 9 variables:  
## $ Pregnancies : int 6 1 8 1 0 5 3 10 2 8 ...  
## $ Glucose : int 148 85 183 89 137 116 78 115 197 125 ...  
## $ BloodPressure : int 72 66 64 66 40 74 50 0 70 96 ...  
## $ SkinThickness : int 35 29 0 23 35 0 32 0 45 0 ...  
## $ Insulin : int 0 0 0 94 168 0 88 0 543 0 ...  
## $ BMI : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...  
## $ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...  
## $ Age : int 50 31 32 21 33 30 26 29 53 54 ...  
## $ Outcome : int 1 0 1 0 1 0 1 0 1 1 ...

categorical\_cols = c('Pregnancies', 'Outcome')  
diabetData[categorical\_cols] = lapply(diabetData[categorical\_cols], as.factor)  
str(diabetData)

## 'data.frame': 768 obs. of 9 variables:  
## $ Pregnancies : Factor w/ 17 levels "0","1","2","3",..: 7 2 9 2 1 6 4 11 3 9 ...  
## $ Glucose : int 148 85 183 89 137 116 78 115 197 125 ...  
## $ BloodPressure : int 72 66 64 66 40 74 50 0 70 96 ...  
## $ SkinThickness : int 35 29 0 23 35 0 32 0 45 0 ...  
## $ Insulin : int 0 0 0 94 168 0 88 0 543 0 ...  
## $ BMI : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...  
## $ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...  
## $ Age : int 50 31 32 21 33 30 26 29 53 54 ...  
## $ Outcome : Factor w/ 2 levels "0","1": 2 1 2 1 2 1 2 1 2 2 ...

#### Visualize the Glucose feature (continuous) using a histogram

p = ggplot(data = diabetData) +  
 geom\_histogram(aes(x = Glucose), binwidth = 10, color = 'steelblue')  
p



mean(diabetData$Glucose)

## [1] 120.8945

median(diabetData$Glucose)

## [1] 117

mean(diabetData$Glucose <= median(diabetData$Glucose))

## [1] 0.5091146

sum(diabetData$Glucose <= median(diabetData$Glucose))

## [1] 391

nrow(diabetData)

## [1] 768

quantile(diabetData$Glucose,0.5) # P(Glucose <+0.5th quantile)=0.5

## 50%   
## 117

quantile(diabetData$Glucose,0.25)

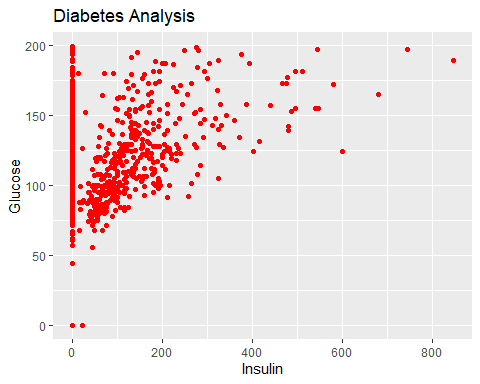
## 25%   
## 99

sum(diabetData$Glucose<=99)/nrow((diabetData))

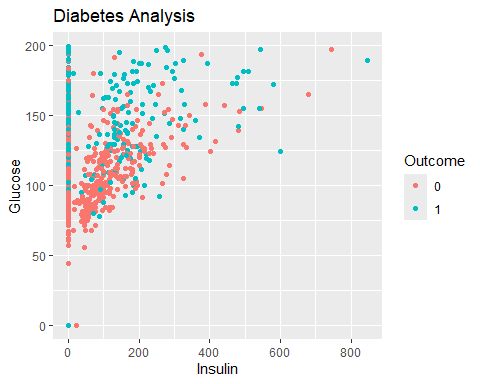
## [1] 0.2565104

#### Scatter plot between Glucose and Insulin (relating two continuous variables)

p = ggplot(data = diabetData) +  
 geom\_point(aes(y = Glucose, x = Insulin), color = 'red') +  
 labs(y = 'Glucose', x = 'Insulin', title = 'Diabetes Analysis')  
p



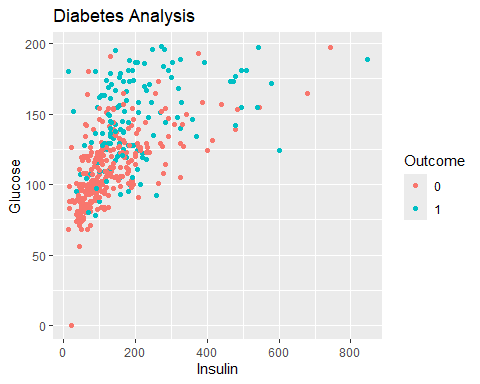
p = ggplot(data = diabetData) +  
 geom\_point(aes(y = Glucose, x = Insulin, color = Outcome)) +  
 labs(y = 'Glucose', x = 'Insulin', title = 'Diabetes Analysis')  
p



data = diabetData %>% filter(Insulin!=0)  
data

## Pregnancies Glucose BloodPressure SkinThickness Insulin BMI  
## 1 1 89 66 23 94 28.1  
## 2 0 137 40 35 168 43.1  
## 3 3 78 50 32 88 31.0  
## 4 2 197 70 45 543 30.5  
## 5 1 189 60 23 846 30.1  
## 6 5 166 72 19 175 25.8  
## 7 0 118 84 47 230 45.8  
## 8 1 103 30 38 83 43.3  
## 9 1 115 70 30 96 34.6  
## 10 3 126 88 41 235 39.3  
## 11 11 143 94 33 146 36.6  
## 12 10 125 70 26 115 31.1  
## 13 1 97 66 15 140 23.2  
## 14 13 145 82 19 110 22.2  
## 15 3 158 76 36 245 31.6  
## 16 3 88 58 11 54 24.8  
## 17 4 103 60 33 192 24.0  
## 18 4 111 72 47 207 37.1  
## 19 3 180 64 25 70 34.0  
## 20 9 171 110 24 240 45.4  
## 21 1 103 80 11 82 19.4  
## 22 1 101 50 15 36 24.2  
## 23 5 88 66 21 23 24.4  
## 24 8 176 90 34 300 33.7  
## 25 7 150 66 42 342 34.7  
## 26 7 187 68 39 304 37.7  
## 27 0 100 88 60 110 46.8  
## 28 0 105 64 41 142 41.5  
## 29 2 141 58 34 128 25.4  
## 30 1 95 66 13 38 19.6  
## 31 4 146 85 27 100 28.9  
## 32 2 100 66 20 90 32.9  
## 33 5 139 64 35 140 28.6  
## 34 4 129 86 20 270 35.1  
## 35 7 83 78 26 71 29.3  
## 36 2 110 74 29 125 32.4  
## 37 2 100 68 25 71 38.5  
## 38 15 136 70 32 110 37.1  
## 39 4 123 80 15 176 32.0  
## 40 7 81 78 40 48 46.7  
## 41 2 142 82 18 64 24.7  
## 42 6 144 72 27 228 33.9  
## 43 1 71 48 18 76 20.4  
## 44 6 93 50 30 64 28.7  
## 45 1 122 90 51 220 49.7  
## 46 1 81 72 18 40 26.6  
## 47 1 126 56 29 152 28.7  
## 48 4 144 58 28 140 29.5  
## 49 3 83 58 31 18 34.3  
## 50 0 95 85 25 36 37.4  
## 51 3 171 72 33 135 33.3  
## 52 8 155 62 26 495 34.0  
## 53 1 89 76 34 37 31.2  
## 54 7 160 54 32 175 30.5  
## 55 4 99 76 15 51 23.2  
## 56 0 162 76 56 100 53.2  
## 57 2 107 74 30 100 33.6  
## 58 1 88 30 42 99 55.0  
## 59 3 120 70 30 135 42.9  
## 60 1 118 58 36 94 33.3  
## 61 1 117 88 24 145 34.5  
## 62 4 173 70 14 168 29.7  
## 63 3 170 64 37 225 34.5  
## 64 2 96 68 13 49 21.1  
## 65 2 125 60 20 140 33.8  
## 66 0 100 70 26 50 30.8  
## 67 0 93 60 25 92 28.7  
## 68 5 105 72 29 325 36.9  
## 69 2 108 52 26 63 32.5  
## 70 4 154 62 31 284 32.8  
## 71 2 106 64 35 119 30.5  
## 72 1 136 74 50 204 37.4  
## 73 9 156 86 28 155 34.3  
## 74 1 153 82 42 485 40.6  
## 75 2 99 52 15 94 24.6  
## 76 1 109 56 21 135 25.2  
## 77 2 88 74 19 53 29.0  
## 78 17 163 72 41 114 40.9  
## 79 7 102 74 40 105 37.2  
## 80 0 114 80 34 285 44.2  
## 81 6 104 74 18 156 29.9  
## 82 3 111 90 12 78 28.4  
## 83 6 134 70 23 130 35.4  
## 84 1 79 60 42 48 43.5  
## 85 2 75 64 24 55 29.7  
## 86 8 179 72 42 130 32.7  
## 87 0 129 110 46 130 67.1  
## 88 0 119 64 18 92 34.9  
## 89 1 0 74 20 23 27.7  
## 90 8 181 68 36 495 30.1  
## 91 1 128 98 41 58 32.0  
## 92 8 109 76 39 114 27.9  
## 93 5 139 80 35 160 31.6  
## 94 9 123 70 44 94 33.1  
## 95 5 158 84 41 210 39.4  
## 96 3 107 62 13 48 22.9  
## 97 4 109 64 44 99 34.8  
## 98 4 148 60 27 318 30.9  
## 99 2 99 70 16 44 20.4  
## 100 6 103 72 32 190 37.7  
## 101 8 196 76 29 280 37.5  
## 102 1 96 64 27 87 33.2  
## 103 0 140 65 26 130 42.6  
## 104 9 112 82 32 175 34.2  
## 105 12 151 70 40 271 41.8  
## 106 5 109 62 41 129 35.8  
## 107 6 125 68 30 120 30.0  
## 108 0 177 60 29 478 34.6  
## 109 7 142 60 33 190 28.8  
## 110 1 100 66 15 56 23.6  
## 111 1 87 78 27 32 34.6  
## 112 4 197 70 39 744 36.7  
## 113 0 117 80 31 53 45.2  
## 114 6 134 80 37 370 46.2  
## 115 1 79 80 25 37 25.4  
## 116 3 74 68 28 45 29.7  
## 117 7 181 84 21 192 35.9  
## 118 4 91 70 32 88 33.1  
## 119 6 119 50 22 176 27.1  
## 120 2 146 76 35 194 38.2  
## 121 0 165 90 33 680 52.3  
## 122 9 124 70 33 402 35.4  
## 123 2 90 80 14 55 24.4  
## 124 12 92 62 7 258 27.6  
## 125 1 193 50 16 375 25.9  
## 126 11 155 76 28 150 33.3  
## 127 3 191 68 15 130 30.9  
## 128 5 96 74 18 67 33.6  
## 129 2 108 62 32 56 25.2  
## 130 1 71 78 50 45 33.2  
## 131 2 100 70 52 57 40.5  
## 132 0 104 64 23 116 27.8  
## 133 2 108 62 10 278 25.3  
## 134 10 129 76 28 122 35.9  
## 135 7 133 88 15 155 32.4  
## 136 7 136 74 26 135 26.0  
## 137 5 155 84 44 545 38.7  
## 138 1 119 86 39 220 45.6  
## 139 4 96 56 17 49 20.8  
## 140 5 108 72 43 75 36.1  
## 141 0 78 88 29 40 36.9  
## 142 0 107 62 30 74 36.6  
## 143 2 128 78 37 182 43.3  
## 144 1 128 48 45 194 40.5  
## 145 6 151 62 31 120 35.5  
## 146 2 146 70 38 360 28.0  
## 147 0 126 84 29 215 30.7  
## 148 14 100 78 25 184 36.6  
## 149 2 144 58 33 135 31.6  
## 150 5 77 82 41 42 35.8  
## 151 2 120 76 37 105 39.7  
## 152 10 161 68 23 132 25.5  
## 153 0 137 68 14 148 24.8  
## 154 0 128 68 19 180 30.5  
## 155 2 124 68 28 205 32.9  
## 156 0 106 70 37 148 39.4  
## 157 2 155 74 17 96 26.6  
## 158 3 113 50 10 85 29.5  
## 159 2 112 68 22 94 34.1  
## 160 3 99 80 11 64 19.3  
## 161 3 115 66 39 140 38.1  
## 162 4 129 60 12 231 27.5  
## 163 13 152 90 33 29 26.8  
## 164 1 157 72 21 168 25.6  
## 165 1 122 64 32 156 35.1  
## 166 2 102 86 36 120 45.5  
## 167 6 105 70 32 68 30.8  
## 168 2 87 58 16 52 32.7  
## 169 1 95 60 18 58 23.9  
## 170 0 165 76 43 255 47.9  
## 171 9 152 78 34 171 34.2  
## 172 1 130 70 13 105 25.9  
## 173 1 95 74 21 73 25.9  
## 174 8 126 88 36 108 38.5  
## 175 1 139 46 19 83 28.7  
## 176 3 99 62 19 74 21.8  
## 177 1 90 62 12 43 27.2  
## 178 1 125 50 40 167 33.3  
## 179 12 88 74 40 54 35.3  
## 180 1 196 76 36 249 36.5  
## 181 5 189 64 33 325 31.2  
## 182 4 147 74 25 293 34.9  
## 183 5 99 54 28 83 34.0  
## 184 3 81 86 16 66 27.5  
## 185 1 133 102 28 140 32.8  
## 186 3 173 82 48 465 38.4  
## 187 0 118 64 23 89 0.0  
## 188 0 84 64 22 66 35.8  
## 189 2 105 58 40 94 34.9  
## 190 2 122 52 43 158 36.2  
## 191 12 140 82 43 325 39.2  
## 192 0 98 82 15 84 25.2  
## 193 1 87 60 37 75 37.2  
## 194 0 93 100 39 72 43.4  
## 195 1 107 72 30 82 30.8  
## 196 1 109 60 8 182 25.4  
## 197 1 90 62 18 59 25.1  
## 198 1 125 70 24 110 24.3  
## 199 1 119 54 13 50 22.3  
## 200 5 144 82 26 285 32.0  
## 201 3 100 68 23 81 31.6  
## 202 1 100 66 29 196 32.0  
## 203 1 131 64 14 415 23.7  
## 204 4 116 72 12 87 22.1  
## 205 2 127 58 24 275 27.7  
## 206 3 96 56 34 115 24.7  
## 207 5 136 84 41 88 35.0  
## 208 2 123 48 32 165 42.1  
## 209 1 172 68 49 579 42.4  
## 210 1 112 72 30 176 34.4  
## 211 1 143 84 23 310 42.4  
## 212 1 143 74 22 61 26.2  
## 213 0 138 60 35 167 34.6  
## 214 3 173 84 33 474 35.7  
## 215 3 129 64 29 115 26.4  
## 216 1 119 88 41 170 45.3  
## 217 2 94 68 18 76 26.0  
## 218 0 102 64 46 78 40.6  
## 219 8 151 78 32 210 42.9  
## 220 4 184 78 39 277 37.0  
## 221 1 181 64 30 180 34.1  
## 222 0 135 94 46 145 40.6  
## 223 1 95 82 25 180 35.0  
## 224 3 89 74 16 85 30.4  
## 225 1 80 74 11 60 30.0  
## 226 2 83 66 23 50 32.2  
## 227 4 117 64 27 120 33.2  
## 228 0 180 78 63 14 59.4  
## 229 1 100 72 12 70 25.3  
## 230 0 95 80 45 92 36.5  
## 231 0 104 64 37 64 33.6  
## 232 0 120 74 18 63 30.5  
## 233 1 82 64 13 95 21.2  
## 234 0 91 68 32 210 39.9  
## 235 2 100 54 28 105 37.8  
## 236 5 86 68 28 71 30.2  
## 237 10 148 84 48 237 37.6  
## 238 9 134 74 33 60 25.9  
## 239 9 120 72 22 56 20.8  
## 240 8 74 70 40 49 35.3  
## 241 0 124 56 13 105 21.8  
## 242 0 74 52 10 36 27.8  
## 243 0 97 64 36 100 36.8  
## 244 6 154 78 41 140 46.1  
## 245 2 105 80 45 191 33.7  
## 246 7 114 76 17 110 23.8  
## 247 8 126 74 38 75 25.9  
## 248 3 158 70 30 328 35.5  
## 249 4 85 58 22 49 27.8  
## 250 0 84 82 31 125 38.2  
## 251 0 135 68 42 250 42.3  
## 252 1 139 62 41 480 40.7  
## 253 0 173 78 32 265 46.5  
## 254 2 83 65 28 66 36.8  
## 255 4 125 70 18 122 28.9  
## 256 2 81 72 15 76 30.1  
## 257 7 195 70 33 145 25.1  
## 258 6 154 74 32 193 29.3  
## 259 2 117 90 19 71 25.2  
## 260 7 94 64 25 79 33.3  
## 261 0 180 90 26 90 36.5  
## 262 1 130 60 23 170 28.6  
## 263 2 84 50 23 76 30.4  
## 264 0 139 62 17 210 22.1  
## 265 3 99 54 19 86 25.6  
## 266 3 163 70 18 105 31.6  
## 267 9 145 88 34 165 30.3  
## 268 6 129 90 7 326 19.6  
## 269 2 68 70 32 66 25.0  
## 270 3 124 80 33 130 33.2  
## 271 1 97 64 19 82 18.2  
## 272 3 116 74 15 105 26.3  
## 273 0 117 66 31 188 30.8  
## 274 2 122 60 18 106 29.8  
## 275 1 86 66 52 65 41.3  
## 276 1 77 56 30 56 33.3  
## 277 0 127 80 37 210 36.3  
## 278 3 129 92 49 155 36.4  
## 279 8 100 74 40 215 39.4  
## 280 3 128 72 25 190 32.4  
## 281 4 84 90 23 56 39.5  
## 282 1 88 78 29 76 32.0  
## 283 8 186 90 35 225 34.5  
## 284 5 187 76 27 207 43.6  
## 285 4 131 68 21 166 33.1  
## 286 1 164 82 43 67 32.8  
## 287 3 84 68 30 106 31.9  
## 288 1 88 62 24 44 29.9  
## 289 1 84 64 23 115 36.9  
## 290 7 124 70 33 215 25.5  
## 291 0 198 66 32 274 41.3  
## 292 1 87 68 34 77 37.6  
## 293 6 99 60 19 54 26.9  
## 294 2 95 54 14 88 26.1  
## 295 1 99 72 30 18 38.6  
## 296 6 92 62 32 126 32.0  
## 297 4 154 72 29 126 31.3  
## 298 0 121 66 30 165 34.3  
## 299 3 111 58 31 44 29.5  
## 300 2 98 60 17 120 34.7  
## 301 1 143 86 30 330 30.1  
## 302 1 119 44 47 63 35.5  
## 303 6 108 44 20 130 24.0  
## 304 8 124 76 24 600 28.7  
## 305 3 176 86 27 156 33.3  
## 306 2 112 78 50 140 39.4  
## 307 2 82 52 22 115 28.5  
## 308 6 123 72 45 230 33.6  
## 309 0 188 82 14 185 32.0  
## 310 1 89 24 19 25 27.8  
## 311 1 109 38 18 120 23.1  
## 312 7 150 78 29 126 35.2  
## 313 1 181 78 42 293 40.0  
## 314 1 92 62 25 41 19.5  
## 315 0 152 82 39 272 41.5  
## 316 1 111 62 13 182 24.0  
## 317 3 106 54 21 158 30.9  
## 318 3 174 58 22 194 32.9  
## 319 7 168 88 42 321 38.2  
## 320 11 138 74 26 144 36.1  
## 321 2 68 62 13 15 20.1  
## 322 2 112 86 42 160 38.4  
## 323 0 94 70 27 115 43.5  
## 324 4 90 88 47 54 37.7  
## 325 0 102 78 40 90 34.5  
## 326 1 128 82 17 183 27.5  
## 327 2 94 76 18 66 31.6  
## 328 7 97 76 32 91 40.9  
## 329 1 100 74 12 46 19.5  
## 330 0 102 86 17 105 29.3  
## 331 3 103 72 30 152 27.6  
## 332 2 157 74 35 440 39.4  
## 333 1 167 74 17 144 23.4  
## 334 0 179 50 36 159 37.8  
## 335 11 136 84 35 130 28.3  
## 336 1 91 54 25 100 25.2  
## 337 1 117 60 23 106 33.8  
## 338 5 123 74 40 77 34.1  
## 339 1 106 70 28 135 34.2  
## 340 2 155 52 27 540 38.7  
## 341 2 101 58 35 90 21.8  
## 342 1 120 80 48 200 38.9  
## 343 3 80 82 31 70 34.2  
## 344 8 167 106 46 231 37.6  
## 345 9 145 80 46 130 37.9  
## 346 1 112 80 45 132 34.8  
## 347 6 98 58 33 190 34.0  
## 348 9 154 78 30 100 30.9  
## 349 6 165 68 26 168 33.6  
## 350 10 68 106 23 49 35.5  
## 351 3 123 100 35 240 57.3  
## 352 2 101 58 17 265 24.2  
## 353 2 56 56 28 45 24.2  
## 354 0 95 64 39 105 44.6  
## 355 2 129 74 26 205 33.2  
## 356 1 140 74 26 180 24.1  
## 357 1 144 82 46 180 46.1  
## 358 2 121 70 32 95 39.1  
## 359 7 129 68 49 125 38.5  
## 360 7 142 90 24 480 30.4  
## 361 3 169 74 19 125 29.9  
## 362 4 127 88 11 155 34.5  
## 363 2 122 76 27 200 35.9  
## 364 4 110 76 20 100 28.4  
## 365 2 127 46 21 335 34.4  
## 366 2 93 64 32 160 38.0  
## 367 3 158 64 13 387 31.2  
## 368 5 126 78 27 22 29.6  
## 369 0 134 58 20 291 26.4  
## 370 7 187 50 33 392 33.9  
## 371 3 173 78 39 185 33.8  
## 372 1 108 60 46 178 35.5  
## 373 1 114 66 36 200 38.1  
## 374 1 149 68 29 127 29.3  
## 375 5 117 86 30 105 39.1  
## 376 1 116 78 29 180 36.1  
## 377 3 130 78 23 79 28.4  
## 378 2 174 88 37 120 44.5  
## 379 2 106 56 27 165 29.0  
## 380 0 126 86 27 120 27.4  
## 381 2 99 60 17 160 36.6  
## 382 11 120 80 37 150 42.3  
## 383 3 102 44 20 94 30.8  
## 384 1 109 58 18 116 28.5  
## 385 13 153 88 37 140 40.6  
## 386 12 100 84 33 105 30.0  
## 387 1 81 74 41 57 46.3  
## 388 3 187 70 22 200 36.4  
## 389 1 121 78 39 74 39.0  
## 390 0 181 88 44 510 43.3  
## 391 1 128 88 39 110 36.5  
## 392 2 88 58 26 16 28.4  
## 393 10 101 76 48 180 32.9  
## 394 5 121 72 23 112 26.2  
## DiabetesPedigreeFunction Age Outcome  
## 1 0.167 21 0  
## 2 2.288 33 1  
## 3 0.248 26 1  
## 4 0.158 53 1  
## 5 0.398 59 1  
## 6 0.587 51 1  
## 7 0.551 31 1  
## 8 0.183 33 0  
## 9 0.529 32 1  
## 10 0.704 27 0  
## 11 0.254 51 1  
## 12 0.205 41 1  
## 13 0.487 22 0  
## 14 0.245 57 0  
## 15 0.851 28 1  
## 16 0.267 22 0  
## 17 0.966 33 0  
## 18 1.390 56 1  
## 19 0.271 26 0  
## 20 0.721 54 1  
## 21 0.491 22 0  
## 22 0.526 26 0  
## 23 0.342 30 0  
## 24 0.467 58 1  
## 25 0.718 42 0  
## 26 0.254 41 1  
## 27 0.962 31 0  
## 28 0.173 22 0  
## 29 0.699 24 0  
## 30 0.334 25 0  
## 31 0.189 27 0  
## 32 0.867 28 1  
## 33 0.411 26 0  
## 34 0.231 23 0  
## 35 0.767 36 0  
## 36 0.698 27 0  
## 37 0.324 26 0  
## 38 0.153 43 1  
## 39 0.443 34 0  
## 40 0.261 42 0  
## 41 0.761 21 0  
## 42 0.255 40 0  
## 43 0.323 22 0  
## 44 0.356 23 0  
## 45 0.325 31 1  
## 46 0.283 24 0  
## 47 0.801 21 0  
## 48 0.287 37 0  
## 49 0.336 25 0  
## 50 0.247 24 1  
## 51 0.199 24 1  
## 52 0.543 46 1  
## 53 0.192 23 0  
## 54 0.588 39 1  
## 55 0.223 21 0  
## 56 0.759 25 1  
## 57 0.404 23 0  
## 58 0.496 26 1  
## 59 0.452 30 0  
## 60 0.261 23 0  
## 61 0.403 40 1  
## 62 0.361 33 1  
## 63 0.356 30 1  
## 64 0.647 26 0  
## 65 0.088 31 0  
## 66 0.597 21 0  
## 67 0.532 22 0  
## 68 0.159 28 0  
## 69 0.318 22 0  
## 70 0.237 23 0  
## 71 1.400 34 0  
## 72 0.399 24 0  
## 73 1.189 42 1  
## 74 0.687 23 0  
## 75 0.637 21 0  
## 76 0.833 23 0  
## 77 0.229 22 0  
## 78 0.817 47 1  
## 79 0.204 45 0  
## 80 0.167 27 0  
## 81 0.722 41 1  
## 82 0.495 29 0  
## 83 0.542 29 1  
## 84 0.678 23 0  
## 85 0.370 33 0  
## 86 0.719 36 1  
## 87 0.319 26 1  
## 88 0.725 23 0  
## 89 0.299 21 0  
## 90 0.615 60 1  
## 91 1.321 33 1  
## 92 0.640 31 1  
## 93 0.361 25 1  
## 94 0.374 40 0  
## 95 0.395 29 1  
## 96 0.678 23 1  
## 97 0.905 26 1  
## 98 0.150 29 1  
## 99 0.235 27 0  
## 100 0.324 55 0  
## 101 0.605 57 1  
## 102 0.289 21 0  
## 103 0.431 24 1  
## 104 0.260 36 1  
## 105 0.742 38 1  
## 106 0.514 25 1  
## 107 0.464 32 0  
## 108 1.072 21 1  
## 109 0.687 61 0  
## 110 0.666 26 0  
## 111 0.101 22 0  
## 112 2.329 31 0  
## 113 0.089 24 0  
## 114 0.238 46 1  
## 115 0.583 22 0  
## 116 0.293 23 0  
## 117 0.586 51 1  
## 118 0.446 22 0  
## 119 1.318 33 1  
## 120 0.329 29 0  
## 121 0.427 23 0  
## 122 0.282 34 0  
## 123 0.249 24 0  
## 124 0.926 44 1  
## 125 0.655 24 0  
## 126 1.353 51 1  
## 127 0.299 34 0  
## 128 0.997 43 0  
## 129 0.128 21 0  
## 130 0.422 21 0  
## 131 0.677 25 0  
## 132 0.454 23 0  
## 133 0.881 22 0  
## 134 0.280 39 0  
## 135 0.262 37 0  
## 136 0.647 51 0  
## 137 0.619 34 0  
## 138 0.808 29 1  
## 139 0.340 26 0  
## 140 0.263 33 0  
## 141 0.434 21 0  
## 142 0.757 25 1  
## 143 1.224 31 1  
## 144 0.613 24 1  
## 145 0.692 28 0  
## 146 0.337 29 1  
## 147 0.520 24 0  
## 148 0.412 46 1  
## 149 0.422 25 1  
## 150 0.156 35 0  
## 151 0.215 29 0  
## 152 0.326 47 1  
## 153 0.143 21 0  
## 154 1.391 25 1  
## 155 0.875 30 1  
## 156 0.605 22 0  
## 157 0.433 27 1  
## 158 0.626 25 0  
## 159 0.315 26 0  
## 160 0.284 30 0  
## 161 0.150 28 0  
## 162 0.527 31 0  
## 163 0.731 43 1  
## 164 0.123 24 0  
## 165 0.692 30 1  
## 166 0.127 23 1  
## 167 0.122 37 0  
## 168 0.166 25 0  
## 169 0.260 22 0  
## 170 0.259 26 0  
## 171 0.893 33 1  
## 172 0.472 22 0  
## 173 0.673 36 0  
## 174 0.349 49 0  
## 175 0.654 22 0  
## 176 0.279 26 0  
## 177 0.580 24 0  
## 178 0.962 28 1  
## 179 0.378 48 0  
## 180 0.875 29 1  
## 181 0.583 29 1  
## 182 0.385 30 0  
## 183 0.499 30 0  
## 184 0.306 22 0  
## 185 0.234 45 1  
## 186 2.137 25 1  
## 187 1.731 21 0  
## 188 0.545 21 0  
## 189 0.225 25 0  
## 190 0.816 28 0  
## 191 0.528 58 1  
## 192 0.299 22 0  
## 193 0.509 22 0  
## 194 1.021 35 0  
## 195 0.821 24 0  
## 196 0.947 21 0  
## 197 1.268 25 0  
## 198 0.221 25 0  
## 199 0.205 24 0  
## 200 0.452 58 1  
## 201 0.949 28 0  
## 202 0.444 42 0  
## 203 0.389 21 0  
## 204 0.463 37 0  
## 205 1.600 25 0  
## 206 0.944 39 0  
## 207 0.286 35 1  
## 208 0.520 26 0  
## 209 0.702 28 1  
## 210 0.528 25 0  
## 211 1.076 22 0  
## 212 0.256 21 0  
## 213 0.534 21 1  
## 214 0.258 22 1  
## 215 0.219 28 1  
## 216 0.507 26 0  
## 217 0.561 21 0  
## 218 0.496 21 0  
## 219 0.516 36 1  
## 220 0.264 31 1  
## 221 0.328 38 1  
## 222 0.284 26 0  
## 223 0.233 43 1  
## 224 0.551 38 0  
## 225 0.527 22 0  
## 226 0.497 22 0  
## 227 0.230 24 0  
## 228 2.420 25 1  
## 229 0.658 28 0  
## 230 0.330 26 0  
## 231 0.510 22 1  
## 232 0.285 26 0  
## 233 0.415 23 0  
## 234 0.381 25 0  
## 235 0.498 24 0  
## 236 0.364 24 0  
## 237 1.001 51 1  
## 238 0.460 81 0  
## 239 0.733 48 0  
## 240 0.705 39 0  
## 241 0.452 21 0  
## 242 0.269 22 0  
## 243 0.600 25 0  
## 244 0.571 27 0  
## 245 0.711 29 1  
## 246 0.466 31 0  
## 247 0.162 39 0  
## 248 0.344 35 1  
## 249 0.306 28 0  
## 250 0.233 23 0  
## 251 0.365 24 1  
## 252 0.536 21 0  
## 253 1.159 58 0  
## 254 0.629 24 0  
## 255 1.144 45 1  
## 256 0.547 25 0  
## 257 0.163 55 1  
## 258 0.839 39 0  
## 259 0.313 21 0  
## 260 0.738 41 0  
## 261 0.314 35 1  
## 262 0.692 21 0  
## 263 0.968 21 0  
## 264 0.207 21 0  
## 265 0.154 24 0  
## 266 0.268 28 1  
## 267 0.771 53 1  
## 268 0.582 60 0  
## 269 0.187 25 0  
## 270 0.305 26 0  
## 271 0.299 21 0  
## 272 0.107 24 0  
## 273 0.493 22 0  
## 274 0.717 22 0  
## 275 0.917 29 0  
## 276 1.251 24 0  
## 277 0.804 23 0  
## 278 0.968 32 1  
## 279 0.661 43 1  
## 280 0.549 27 1  
## 281 0.159 25 0  
## 282 0.365 29 0  
## 283 0.423 37 1  
## 284 1.034 53 1  
## 285 0.160 28 0  
## 286 0.341 50 0  
## 287 0.591 25 0  
## 288 0.422 23 0  
## 289 0.471 28 0  
## 290 0.161 37 0  
## 291 0.502 28 1  
## 292 0.401 24 0  
## 293 0.497 32 0  
## 294 0.748 22 0  
## 295 0.412 21 0  
## 296 0.085 46 0  
## 297 0.338 37 0  
## 298 0.203 33 1  
## 299 0.430 22 0  
## 300 0.198 22 0  
## 301 0.892 23 0  
## 302 0.280 25 0  
## 303 0.813 35 0  
## 304 0.687 52 1  
## 305 1.154 52 1  
## 306 0.175 24 0  
## 307 1.699 25 0  
## 308 0.733 34 0  
## 309 0.682 22 1  
## 310 0.559 21 0  
## 311 0.407 26 0  
## 312 0.692 54 1  
## 313 1.258 22 1  
## 314 0.482 25 0  
## 315 0.270 27 0  
## 316 0.138 23 0  
## 317 0.292 24 0  
## 318 0.593 36 1  
## 319 0.787 40 1  
## 320 0.557 50 1  
## 321 0.257 23 0  
## 322 0.246 28 0  
## 323 0.347 21 0  
## 324 0.362 29 0  
## 325 0.238 24 0  
## 326 0.115 22 0  
## 327 0.649 23 0  
## 328 0.871 32 1  
## 329 0.149 28 0  
## 330 0.695 27 0  
## 331 0.730 27 0  
## 332 0.134 30 0  
## 333 0.447 33 1  
## 334 0.455 22 1  
## 335 0.260 42 1  
## 336 0.234 23 0  
## 337 0.466 27 0  
## 338 0.269 28 0  
## 339 0.142 22 0  
## 340 0.240 25 1  
## 341 0.155 22 0  
## 342 1.162 41 0  
## 343 1.292 27 1  
## 344 0.165 43 1  
## 345 0.637 40 1  
## 346 0.217 24 0  
## 347 0.430 43 0  
## 348 0.164 45 0  
## 349 0.631 49 0  
## 350 0.285 47 0  
## 351 0.880 22 0  
## 352 0.614 23 0  
## 353 0.332 22 0  
## 354 0.366 22 0  
## 355 0.591 25 0  
## 356 0.828 23 0  
## 357 0.335 46 1  
## 358 0.886 23 0  
## 359 0.439 43 1  
## 360 0.128 43 1  
## 361 0.268 31 1  
## 362 0.598 28 0  
## 363 0.483 26 0  
## 364 0.118 27 0  
## 365 0.176 22 0  
## 366 0.674 23 1  
## 367 0.295 24 0  
## 368 0.439 40 0  
## 369 0.352 21 0  
## 370 0.826 34 1  
## 371 0.970 31 1  
## 372 0.415 24 0  
## 373 0.289 21 0  
## 374 0.349 42 1  
## 375 0.251 42 0  
## 376 0.496 25 0  
## 377 0.323 34 1  
## 378 0.646 24 1  
## 379 0.426 22 0  
## 380 0.515 21 0  
## 381 0.453 21 0  
## 382 0.785 48 1  
## 383 0.400 26 0  
## 384 0.219 22 0  
## 385 1.174 39 0  
## 386 0.488 46 0  
## 387 1.096 32 0  
## 388 0.408 36 1  
## 389 0.261 28 0  
## 390 0.222 26 1  
## 391 1.057 37 1  
## 392 0.766 22 0  
## 393 0.171 63 0  
## 394 0.245 30 0

p = ggplot(data = diabetData %>% filter(Insulin!=0)) +  
 geom\_point(aes(y = Glucose, x = Insulin, color = Outcome)) +  
 labs(y = 'Glucose', x = 'Insulin', title = 'Diabetes Analysis')  
p



diabetData %>% select(Glucose)

## Glucose  
## 1 148  
## 2 85  
## 3 183  
## 4 89  
## 5 137  
## 6 116  
## 7 78  
## 8 115  
## 9 197  
## 10 125  
## 11 110  
## 12 168  
## 13 139  
## 14 189  
## 15 166  
## 16 100  
## 17 118  
## 18 107  
## 19 103  
## 20 115  
## 21 126  
## 22 99  
## 23 196  
## 24 119  
## 25 143  
## 26 125  
## 27 147  
## 28 97  
## 29 145  
## 30 117  
## 31 109  
## 32 158  
## 33 88  
## 34 92  
## 35 122  
## 36 103  
## 37 138  
## 38 102  
## 39 90  
## 40 111  
## 41 180  
## 42 133  
## 43 106  
## 44 171  
## 45 159  
## 46 180  
## 47 146  
## 48 71  
## 49 103  
## 50 105  
## 51 103  
## 52 101  
## 53 88  
## 54 176  
## 55 150  
## 56 73  
## 57 187  
## 58 100  
## 59 146  
## 60 105  
## 61 84  
## 62 133  
## 63 44  
## 64 141  
## 65 114  
## 66 99  
## 67 109  
## 68 109  
## 69 95  
## 70 146  
## 71 100  
## 72 139  
## 73 126  
## 74 129  
## 75 79  
## 76 0  
## 77 62  
## 78 95  
## 79 131  
## 80 112  
## 81 113  
## 82 74  
## 83 83  
## 84 101  
## 85 137  
## 86 110  
## 87 106  
## 88 100  
## 89 136  
## 90 107  
## 91 80  
## 92 123  
## 93 81  
## 94 134  
## 95 142  
## 96 144  
## 97 92  
## 98 71  
## 99 93  
## 100 122  
## 101 163  
## 102 151  
## 103 125  
## 104 81  
## 105 85  
## 106 126  
## 107 96  
## 108 144  
## 109 83  
## 110 95  
## 111 171  
## 112 155  
## 113 89  
## 114 76  
## 115 160  
## 116 146  
## 117 124  
## 118 78  
## 119 97  
## 120 99  
## 121 162  
## 122 111  
## 123 107  
## 124 132  
## 125 113  
## 126 88  
## 127 120  
## 128 118  
## 129 117  
## 130 105  
## 131 173  
## 132 122  
## 133 170  
## 134 84  
## 135 96  
## 136 125  
## 137 100  
## 138 93  
## 139 129  
## 140 105  
## 141 128  
## 142 106  
## 143 108  
## 144 108  
## 145 154  
## 146 102  
## 147 57  
## 148 106  
## 149 147  
## 150 90  
## 151 136  
## 152 114  
## 153 156  
## 154 153  
## 155 188  
## 156 152  
## 157 99  
## 158 109  
## 159 88  
## 160 163  
## 161 151  
## 162 102  
## 163 114  
## 164 100  
## 165 131  
## 166 104  
## 167 148  
## 168 120  
## 169 110  
## 170 111  
## 171 102  
## 172 134  
## 173 87  
## 174 79  
## 175 75  
## 176 179  
## 177 85  
## 178 129  
## 179 143  
## 180 130  
## 181 87  
## 182 119  
## 183 0  
## 184 73  
## 185 141  
## 186 194  
## 187 181  
## 188 128  
## 189 109  
## 190 139  
## 191 111  
## 192 123  
## 193 159  
## 194 135  
## 195 85  
## 196 158  
## 197 105  
## 198 107  
## 199 109  
## 200 148  
## 201 113  
## 202 138  
## 203 108  
## 204 99  
## 205 103  
## 206 111  
## 207 196  
## 208 162  
## 209 96  
## 210 184  
## 211 81  
## 212 147  
## 213 179  
## 214 140  
## 215 112  
## 216 151  
## 217 109  
## 218 125  
## 219 85  
## 220 112  
## 221 177  
## 222 158  
## 223 119  
## 224 142  
## 225 100  
## 226 87  
## 227 101  
## 228 162  
## 229 197  
## 230 117  
## 231 142  
## 232 134  
## 233 79  
## 234 122  
## 235 74  
## 236 171  
## 237 181  
## 238 179  
## 239 164  
## 240 104  
## 241 91  
## 242 91  
## 243 139  
## 244 119  
## 245 146  
## 246 184  
## 247 122  
## 248 165  
## 249 124  
## 250 111  
## 251 106  
## 252 129  
## 253 90  
## 254 86  
## 255 92  
## 256 113  
## 257 111  
## 258 114  
## 259 193  
## 260 155  
## 261 191  
## 262 141  
## 263 95  
## 264 142  
## 265 123  
## 266 96  
## 267 138  
## 268 128  
## 269 102  
## 270 146  
## 271 101  
## 272 108  
## 273 122  
## 274 71  
## 275 106  
## 276 100  
## 277 106  
## 278 104  
## 279 114  
## 280 108  
## 281 146  
## 282 129  
## 283 133  
## 284 161  
## 285 108  
## 286 136  
## 287 155  
## 288 119  
## 289 96  
## 290 108  
## 291 78  
## 292 107  
## 293 128  
## 294 128  
## 295 161  
## 296 151  
## 297 146  
## 298 126  
## 299 100  
## 300 112  
## 301 167  
## 302 144  
## 303 77  
## 304 115  
## 305 150  
## 306 120  
## 307 161  
## 308 137  
## 309 128  
## 310 124  
## 311 80  
## 312 106  
## 313 155  
## 314 113  
## 315 109  
## 316 112  
## 317 99  
## 318 182  
## 319 115  
## 320 194  
## 321 129  
## 322 112  
## 323 124  
## 324 152  
## 325 112  
## 326 157  
## 327 122  
## 328 179  
## 329 102  
## 330 105  
## 331 118  
## 332 87  
## 333 180  
## 334 106  
## 335 95  
## 336 165  
## 337 117  
## 338 115  
## 339 152  
## 340 178  
## 341 130  
## 342 95  
## 343 0  
## 344 122  
## 345 95  
## 346 126  
## 347 139  
## 348 116  
## 349 99  
## 350 0  
## 351 92  
## 352 137  
## 353 61  
## 354 90  
## 355 90  
## 356 165  
## 357 125  
## 358 129  
## 359 88  
## 360 196  
## 361 189  
## 362 158  
## 363 103  
## 364 146  
## 365 147  
## 366 99  
## 367 124  
## 368 101  
## 369 81  
## 370 133  
## 371 173  
## 372 118  
## 373 84  
## 374 105  
## 375 122  
## 376 140  
## 377 98  
## 378 87  
## 379 156  
## 380 93  
## 381 107  
## 382 105  
## 383 109  
## 384 90  
## 385 125  
## 386 119  
## 387 116  
## 388 105  
## 389 144  
## 390 100  
## 391 100  
## 392 166  
## 393 131  
## 394 116  
## 395 158  
## 396 127  
## 397 96  
## 398 131  
## 399 82  
## 400 193  
## 401 95  
## 402 137  
## 403 136  
## 404 72  
## 405 168  
## 406 123  
## 407 115  
## 408 101  
## 409 197  
## 410 172  
## 411 102  
## 412 112  
## 413 143  
## 414 143  
## 415 138  
## 416 173  
## 417 97  
## 418 144  
## 419 83  
## 420 129  
## 421 119  
## 422 94  
## 423 102  
## 424 115  
## 425 151  
## 426 184  
## 427 94  
## 428 181  
## 429 135  
## 430 95  
## 431 99  
## 432 89  
## 433 80  
## 434 139  
## 435 90  
## 436 141  
## 437 140  
## 438 147  
## 439 97  
## 440 107  
## 441 189  
## 442 83  
## 443 117  
## 444 108  
## 445 117  
## 446 180  
## 447 100  
## 448 95  
## 449 104  
## 450 120  
## 451 82  
## 452 134  
## 453 91  
## 454 119  
## 455 100  
## 456 175  
## 457 135  
## 458 86  
## 459 148  
## 460 134  
## 461 120  
## 462 71  
## 463 74  
## 464 88  
## 465 115  
## 466 124  
## 467 74  
## 468 97  
## 469 120  
## 470 154  
## 471 144  
## 472 137  
## 473 119  
## 474 136  
## 475 114  
## 476 137  
## 477 105  
## 478 114  
## 479 126  
## 480 132  
## 481 158  
## 482 123  
## 483 85  
## 484 84  
## 485 145  
## 486 135  
## 487 139  
## 488 173  
## 489 99  
## 490 194  
## 491 83  
## 492 89  
## 493 99  
## 494 125  
## 495 80  
## 496 166  
## 497 110  
## 498 81  
## 499 195  
## 500 154  
## 501 117  
## 502 84  
## 503 0  
## 504 94  
## 505 96  
## 506 75  
## 507 180  
## 508 130  
## 509 84  
## 510 120  
## 511 84  
## 512 139  
## 513 91  
## 514 91  
## 515 99  
## 516 163  
## 517 145  
## 518 125  
## 519 76  
## 520 129  
## 521 68  
## 522 124  
## 523 114  
## 524 130  
## 525 125  
## 526 87  
## 527 97  
## 528 116  
## 529 117  
## 530 111  
## 531 122  
## 532 107  
## 533 86  
## 534 91  
## 535 77  
## 536 132  
## 537 105  
## 538 57  
## 539 127  
## 540 129  
## 541 100  
## 542 128  
## 543 90  
## 544 84  
## 545 88  
## 546 186  
## 547 187  
## 548 131  
## 549 164  
## 550 189  
## 551 116  
## 552 84  
## 553 114  
## 554 88  
## 555 84  
## 556 124  
## 557 97  
## 558 110  
## 559 103  
## 560 85  
## 561 125  
## 562 198  
## 563 87  
## 564 99  
## 565 91  
## 566 95  
## 567 99  
## 568 92  
## 569 154  
## 570 121  
## 571 78  
## 572 130  
## 573 111  
## 574 98  
## 575 143  
## 576 119  
## 577 108  
## 578 118  
## 579 133  
## 580 197  
## 581 151  
## 582 109  
## 583 121  
## 584 100  
## 585 124  
## 586 93  
## 587 143  
## 588 103  
## 589 176  
## 590 73  
## 591 111  
## 592 112  
## 593 132  
## 594 82  
## 595 123  
## 596 188  
## 597 67  
## 598 89  
## 599 173  
## 600 109  
## 601 108  
## 602 96  
## 603 124  
## 604 150  
## 605 183  
## 606 124  
## 607 181  
## 608 92  
## 609 152  
## 610 111  
## 611 106  
## 612 174  
## 613 168  
## 614 105  
## 615 138  
## 616 106  
## 617 117  
## 618 68  
## 619 112  
## 620 119  
## 621 112  
## 622 92  
## 623 183  
## 624 94  
## 625 108  
## 626 90  
## 627 125  
## 628 132  
## 629 128  
## 630 94  
## 631 114  
## 632 102  
## 633 111  
## 634 128  
## 635 92  
## 636 104  
## 637 104  
## 638 94  
## 639 97  
## 640 100  
## 641 102  
## 642 128  
## 643 147  
## 644 90  
## 645 103  
## 646 157  
## 647 167  
## 648 179  
## 649 136  
## 650 107  
## 651 91  
## 652 117  
## 653 123  
## 654 120  
## 655 106  
## 656 155  
## 657 101  
## 658 120  
## 659 127  
## 660 80  
## 661 162  
## 662 199  
## 663 167  
## 664 145  
## 665 115  
## 666 112  
## 667 145  
## 668 111  
## 669 98  
## 670 154  
## 671 165  
## 672 99  
## 673 68  
## 674 123  
## 675 91  
## 676 195  
## 677 156  
## 678 93  
## 679 121  
## 680 101  
## 681 56  
## 682 162  
## 683 95  
## 684 125  
## 685 136  
## 686 129  
## 687 130  
## 688 107  
## 689 140  
## 690 144  
## 691 107  
## 692 158  
## 693 121  
## 694 129  
## 695 90  
## 696 142  
## 697 169  
## 698 99  
## 699 127  
## 700 118  
## 701 122  
## 702 125  
## 703 168  
## 704 129  
## 705 110  
## 706 80  
## 707 115  
## 708 127  
## 709 164  
## 710 93  
## 711 158  
## 712 126  
## 713 129  
## 714 134  
## 715 102  
## 716 187  
## 717 173  
## 718 94  
## 719 108  
## 720 97  
## 721 83  
## 722 114  
## 723 149  
## 724 117  
## 725 111  
## 726 112  
## 727 116  
## 728 141  
## 729 175  
## 730 92  
## 731 130  
## 732 120  
## 733 174  
## 734 106  
## 735 105  
## 736 95  
## 737 126  
## 738 65  
## 739 99  
## 740 102  
## 741 120  
## 742 102  
## 743 109  
## 744 140  
## 745 153  
## 746 100  
## 747 147  
## 748 81  
## 749 187  
## 750 162  
## 751 136  
## 752 121  
## 753 108  
## 754 181  
## 755 154  
## 756 128  
## 757 137  
## 758 123  
## 759 106  
## 760 190  
## 761 88  
## 762 170  
## 763 89  
## 764 101  
## 765 122  
## 766 121  
## 767 126  
## 768 93

Glucose\_median=median(diabetData$Glucose) # 50 th percentile or 0.5th quantile  
print(Glucose\_median)

## [1] 117

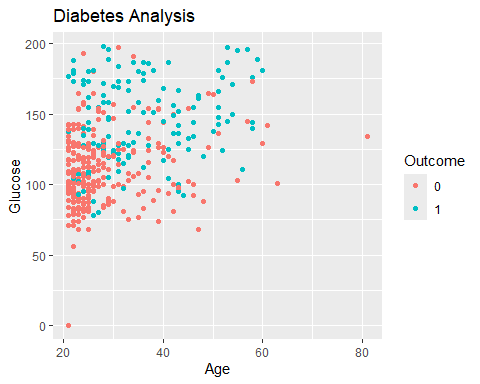
nrow(diabetData %>% filter(Glucose <= Glucose\_median) %>% select(Glucose))/nrow(diabetData)

## [1] 0.5091146

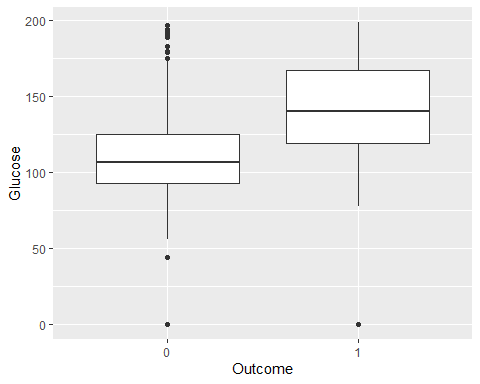
mean(diabetData$Glucose <=Glucose\_median)

## [1] 0.5091146

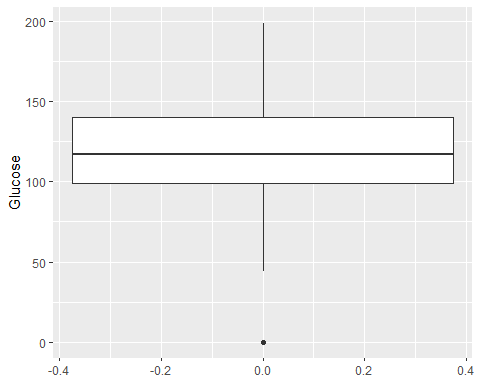
p = ggplot(data = diabetData %>% filter(Insulin!=0)) +  
 geom\_point(aes(x = Age, y = Glucose, color = Outcome)) +  
 labs(y = 'Glucose', x = 'Age', title = 'Diabetes Analysis')  
p



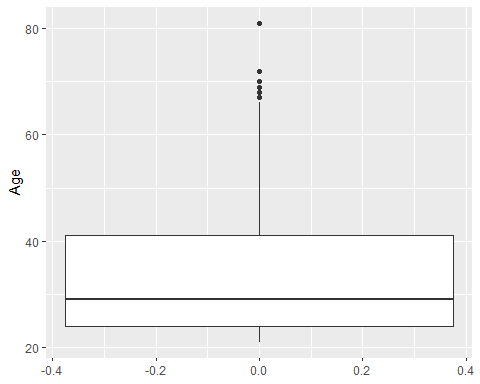
p = ggplot(data = diabetData) +  
 geom\_boxplot(aes(x = Outcome, y = Glucose))  
p



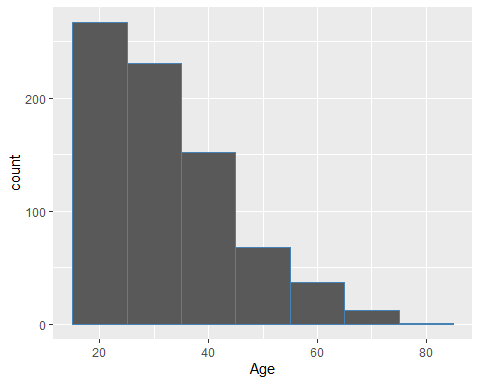
p = ggplot(data = diabetData) +  
 geom\_boxplot(aes(y = Glucose))  
p



p = ggplot(data = diabetData) +  
 geom\_boxplot(aes(y = Age))  
p



p = ggplot(data = diabetData) +  
 geom\_histogram(aes(x = Age), binwidth = 10, color = 'steelblue')  
p



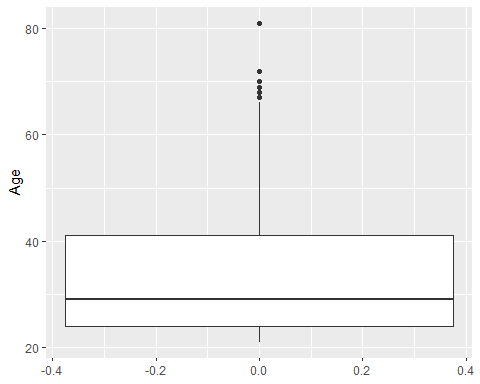
mean(diabetData$Age)

## [1] 33.24089

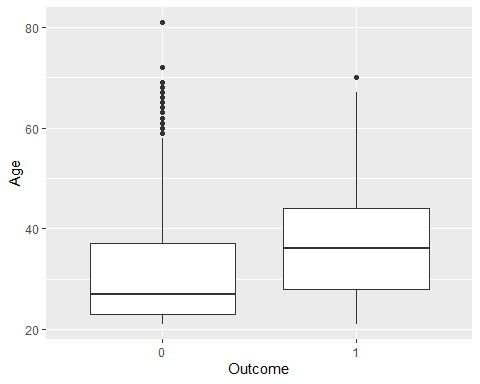
median(diabetData$Age)

## [1] 29

p = ggplot(data = diabetData) +  
 geom\_boxplot(aes(y = Age))  
p



p = ggplot(data = diabetData) +  
 geom\_boxplot(aes(y = Age , x= Outcome))  
p



#p = ggplot(data = diabetData) +  
 #geom\_boxplot(aes(y = Outcome))  
#p

p = ggplot(data = diabetData %>% filter(Insulin!=0)) +  
 geom\_boxplot(aes(y = Age , x= Outcome))  
p

