Useful Visualizations

Jasmine Horan, Shukry Zablah

06 December, 2018

Contents

Imports 1
Load Data 1
Visualizations 1
Percentages of Diabetes in Binned Groups 1
Percentages of Diabetes in Binned BMI Groups 2
Percentages of Diabetes in Binned Glucose Concentration Groups 3
Percentages of Diabetes in Binned Diabetes Pedigree Function 4

Imports

```
library(dplyr)
library(tidyr)
library(ggplot2)
```

Load Data

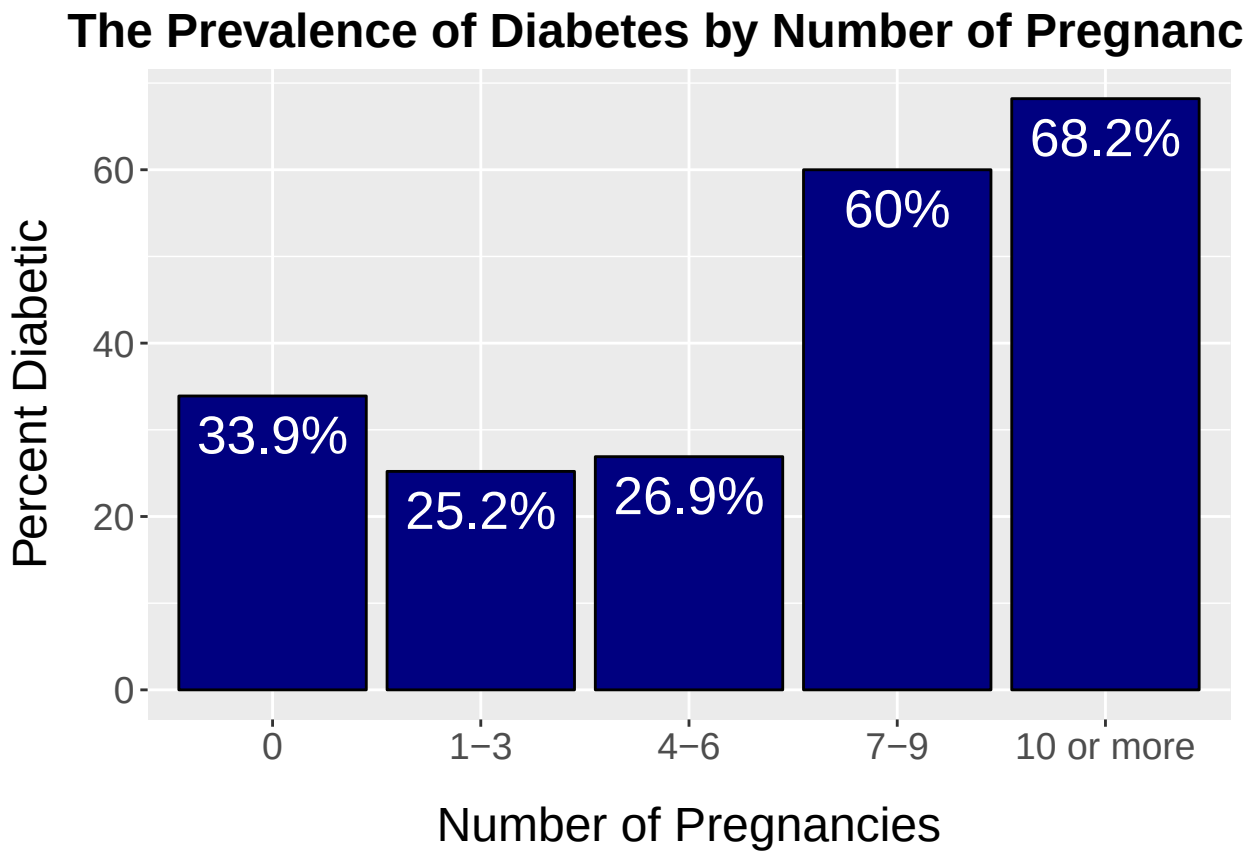
```
PIMA <- readRDS(file = "../data/PIMA_noNAs.Rds")
```

Visualizations

Percentages of Diabetes in Binned Groups

```
PIMA %>%
  mutate(pregnancies = cut(pregnancies,
                           breaks = c(-.5,.5,3.5,6.5,9.5, 50),
                           labels = c("0", "1-3", "4-6", "7-9", "10 or more"))) %>%
  group_by(pregnancies, hasDiabetes) %>%
  summarize(count = n()) %>%
  spread(hasDiabetes, count) %>%
  mutate(percentage = round(100*(`1`/(`0`+`1`)),1)) %>%
  ggplot(aes(x = pregnancies, y = percentage, label = paste0(percentage, "%"))) +
    geom_bar(stat = "identity", fill = "navy", color = "black") +
    geom_text(vjust=1.5, size = 7, color="white") +
    labs(x = "Number of Pregnancies",
         y = "Percent Diabetic",
         title = "The Prevalence of Diabetes by Number of Pregnancies") +
    scale_x_discrete(limits = c("0", "1-3", "4-6", "7-9", "10 or more")) +
    theme(axis.text.x = element_text(size = 14),
          axis.title.x = element_text(size = 18, margin = margin(t = 15)),
          axis.text.y = element_text(size = 14),
```

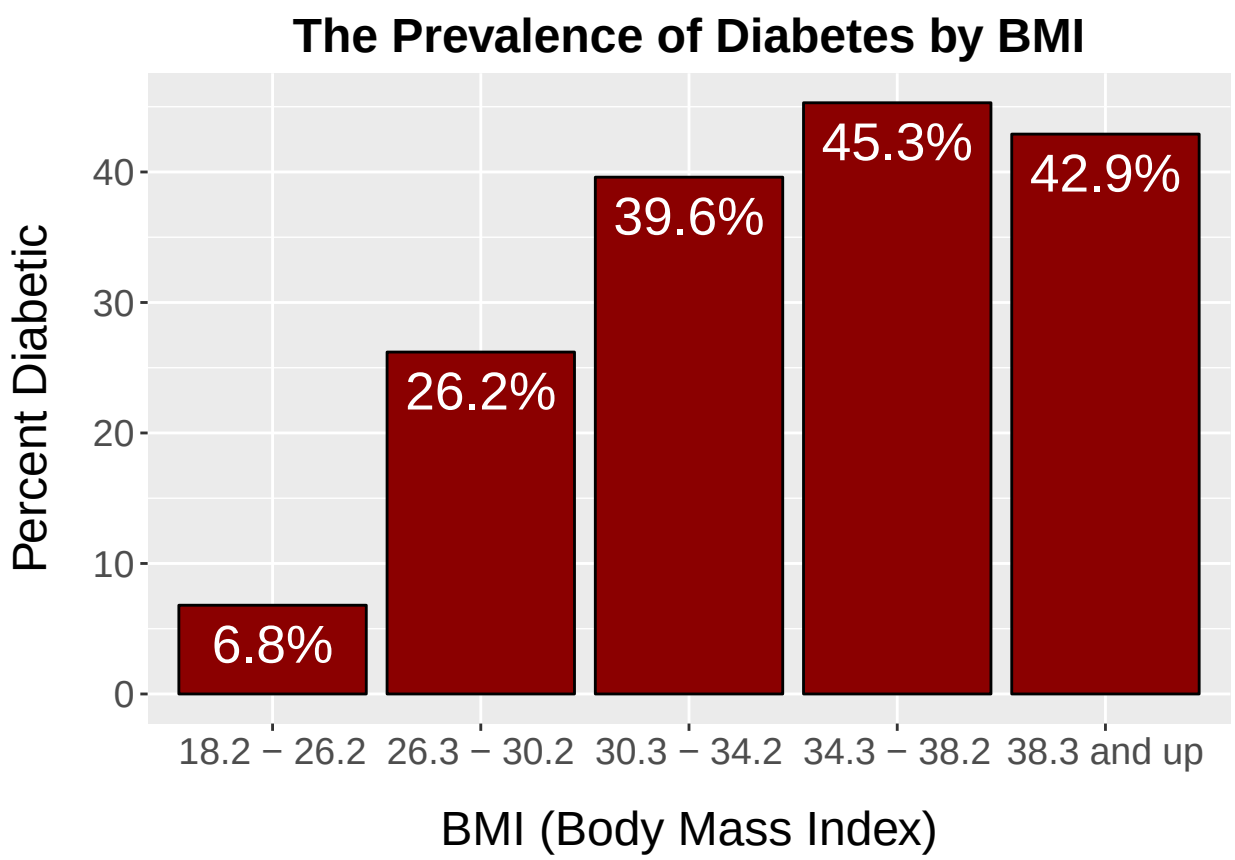
```
        axis.title.y = element_text(size = 18, margin = margin(r = 15)),
        plot.title = element_text(hjust = 0.5, face = "bold", size = 18))
```



Percentages of Diabetes in Binned BMI Groups

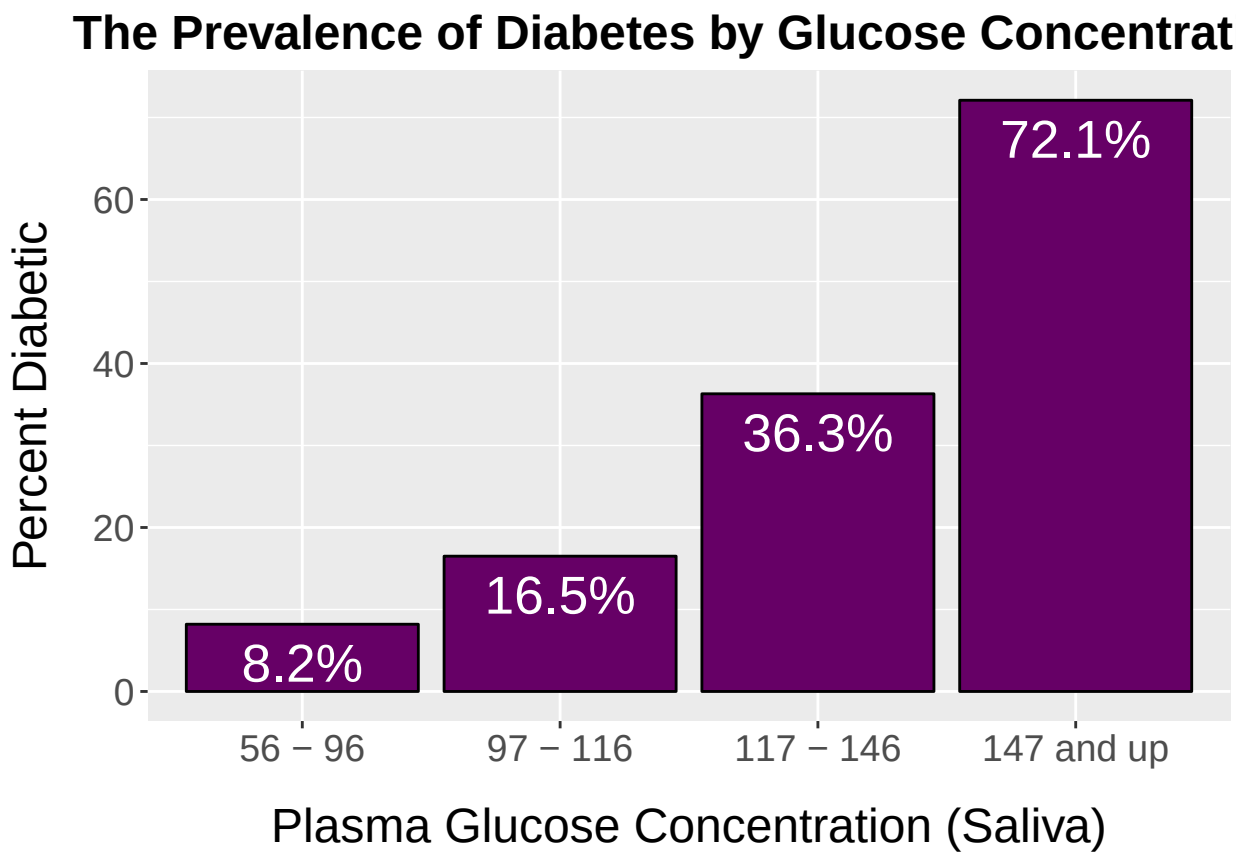
```
PIMA %>%
  mutate(bmi = cut(bmi,
                           breaks = c(0,26.2, 30.2, 34.2, 38.2, 100),
                           labels = c("18.2 - 26.2", "26.3 - 30.2", "30.3 - 34.2", "34.3 - 38.2", "38.3
  group_by(bmi, hasDiabetes) %>%
  summarize(count = n()) %>%
  spread(hasDiabetes, count) %>%
  mutate(percentage = round(100*(`1`/(`0`+`1`)),1)) %>%
  ggplot(aes(x = bmi, y = percentage, label = paste0(percentage, "%"))) +
    geom_bar(stat = "identity", fill = "dark red", color = "black") +
    geom_text(vjust=1.5, size = 7, color="white") +
    labs(x = "BMI (Body Mass Index)",
         y = "Percent Diabetic",
         title = "The Prevalence of Diabetes by BMI") +
    scale_x_discrete(limits = c("18.2 - 26.2", "26.3 - 30.2", "30.3 - 34.2", "34.3 - 38.2", "38.3 and up
    theme(axis.text.x = element_text(size = 14),
          axis.title.x = element_text(size = 18, margin = margin(t = 15)),
          axis.text.y = element_text(size = 14),
          axis.title.y = element_text(size = 18, margin = margin(r = 15)),
```

```
        plot.title = element_text(hjust = 0.5, face = "bold", size = 18))
```



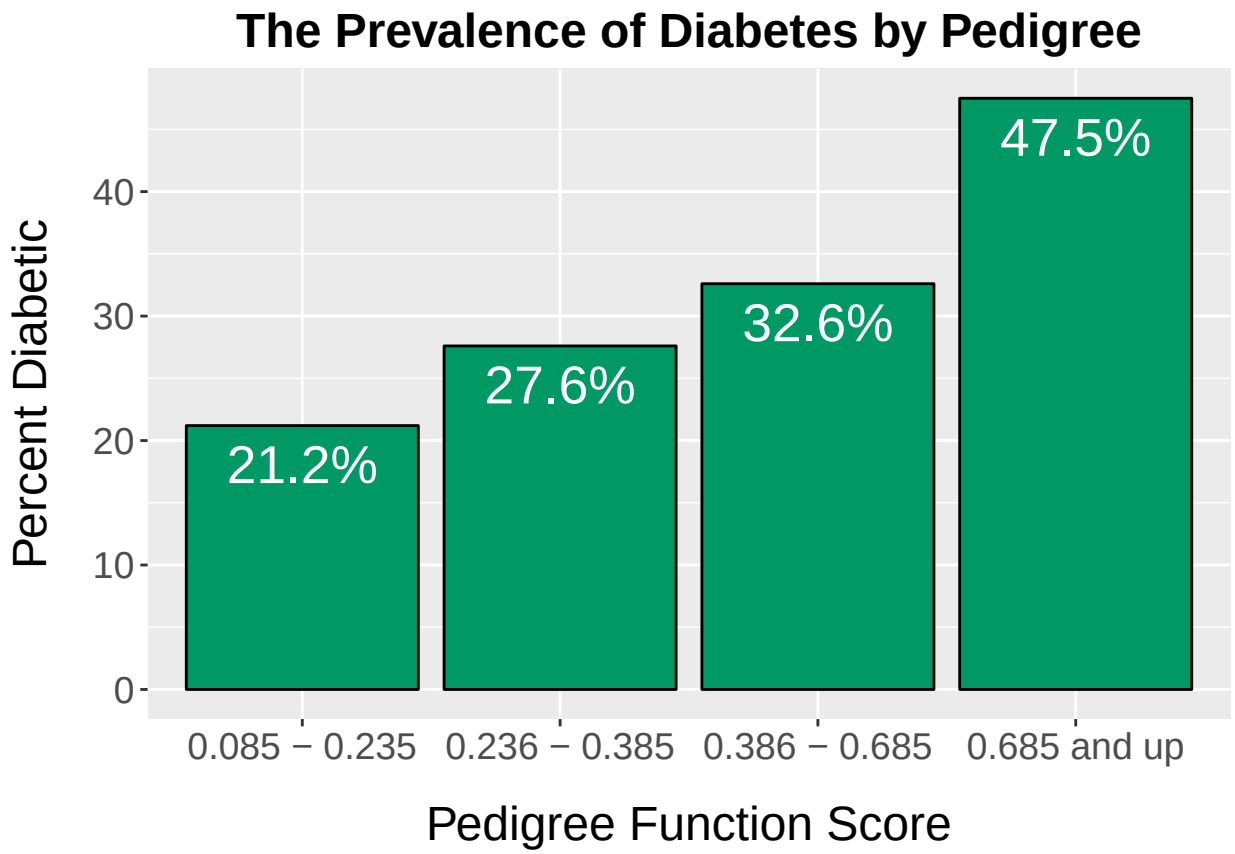
Percentages of Diabetes in Binned Glucose Concentration Groups

```
PIMA %>%
  mutate(glucoseConcentration = cut(glucoseConcentration,
                          breaks = c(50, 96, 116, 146, 200),
                          labels = c("56 - 96", "97 - 116", "117 - 146", "147 and up"))) %>%
  group_by(glucoseConcentration, hasDiabetes) %>%
  summarize(count = n()) %>%
  spread(hasDiabetes, count) %>%
  mutate(percentage = round(100*(`1`/(`0`+`1`)),1)) %>%
  ggplot(aes(x = glucoseConcentration, y = percentage, label = paste0(percentage, "%"))) +
    geom_bar(stat = "identity", fill = "#660066", color = "black") +
    geom_text(vjust=1.5, size = 7, color="white") +
    labs(x = "Plasma Glucose Concentration (Saliva)",
         y = "Percent Diabetic",
         title = "The Prevalence of Diabetes by Glucose Concentration") +
    scale_x_discrete(limits = c("56 - 96", "97 - 116", "117 - 146", "147 and up")) +
    theme(axis.text.x = element_text(size = 14),
          axis.title.x = element_text(size = 18, margin = margin(t = 15)),
          axis.text.y = element_text(size = 14),
          axis.title.y = element_text(size = 18, margin = margin(r = 15)),
          plot.title = element_text(hjust = 0.5, face = "bold", size = 18))
```



Percentages of Diabetes in Binned Diabetes Pedigree Function

```
PIMA %>%
  mutate(diabetesPedigreeFunction = cut(diabetesPedigreeFunction,
                           breaks = c(0,0.235, 0.385, 0.685, 3),
                           labels = c("0.085 - 0.235", "0.236 - 0.385", "0.386 - 0.685", "0.685 and up"))
  group_by(diabetesPedigreeFunction, hasDiabetes) %>%
  summarize(count = n()) %>%
  spread(hasDiabetes, count) %>%
  mutate(percentage = round(100*(`1`/(`0`+`1`)),1)) %>%
  ggplot(aes(x = diabetesPedigreeFunction, y = percentage, label = paste0(percentage, "%"))) +
    geom_bar(stat = "identity", fill = "#009966", color = "black") +
    geom_text(vjust=1.5, size = 7, color="white") +
    labs(x = "Pedigree Function Score",
         y = "Percent Diabetic",
         title = "The Prevalence of Diabetes by Pedigree") +
    scale_x_discrete(limits = c("0.085 - 0.235", "0.236 - 0.385", "0.386 - 0.685", "0.685 and up")) +
    theme(axis.text.x = element_text(size = 14),
          axis.title.x = element_text(size = 18, margin = margin(t = 15)),
          axis.text.y = element_text(size = 14),
          axis.title.y = element_text(size = 18, margin = margin(r = 15)),
          plot.title = element_text(hjust = 0.5, face = "bold", size = 18))
```


The Prevalence of Diabetes by Pedigree
Percent Diabetic
40
30
20
10
0
21.2%
27.6%
32.6%
47.5%
0.085 − 0.235
0.236 − 0.385
0.386 − 0.685
0.685 and up
Pedigree Function Score