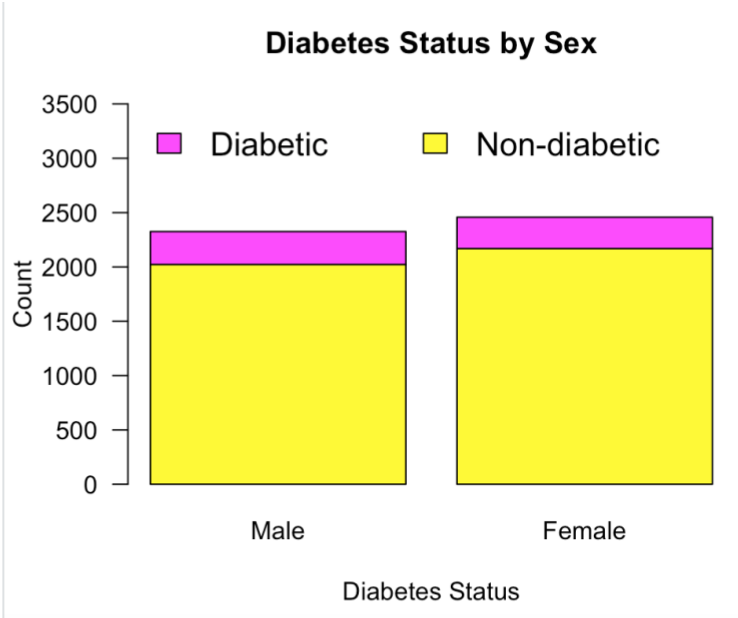
Stephanie Mecham
EPID 674 Section 2

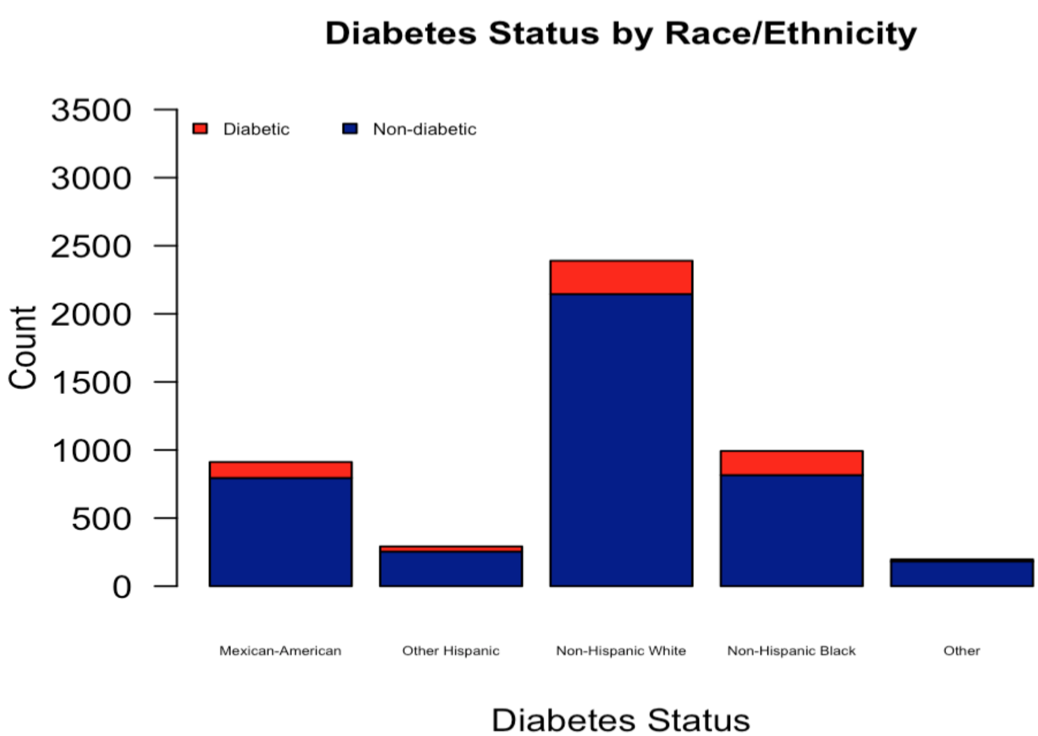
Exercise 5A

Problem 1.

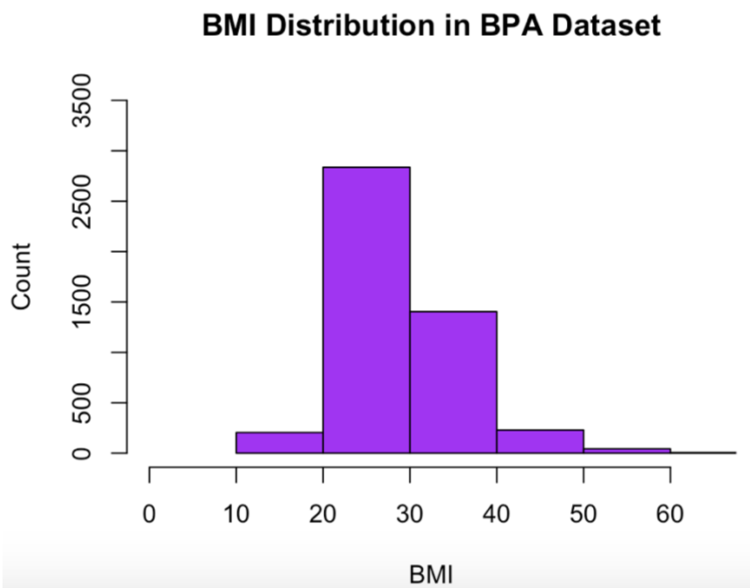
T2DM x Sex



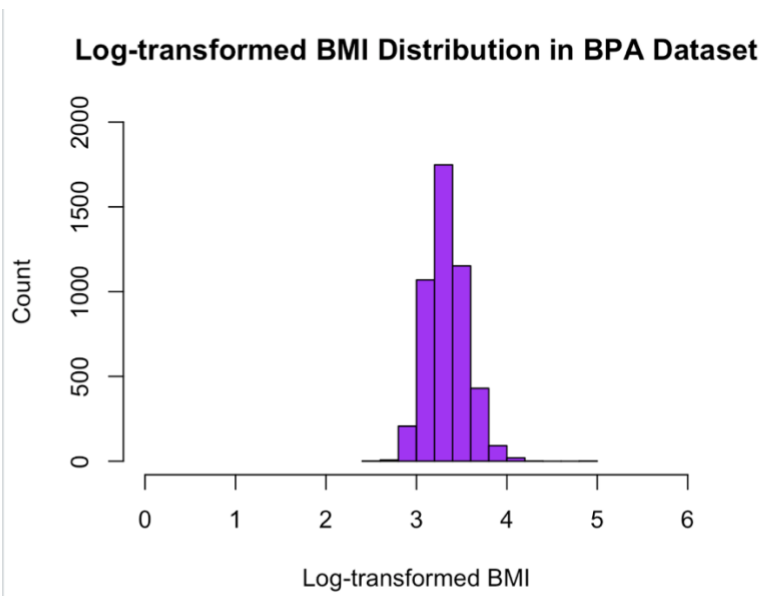
T2DM x Race/Ethnicity



Problem 2.

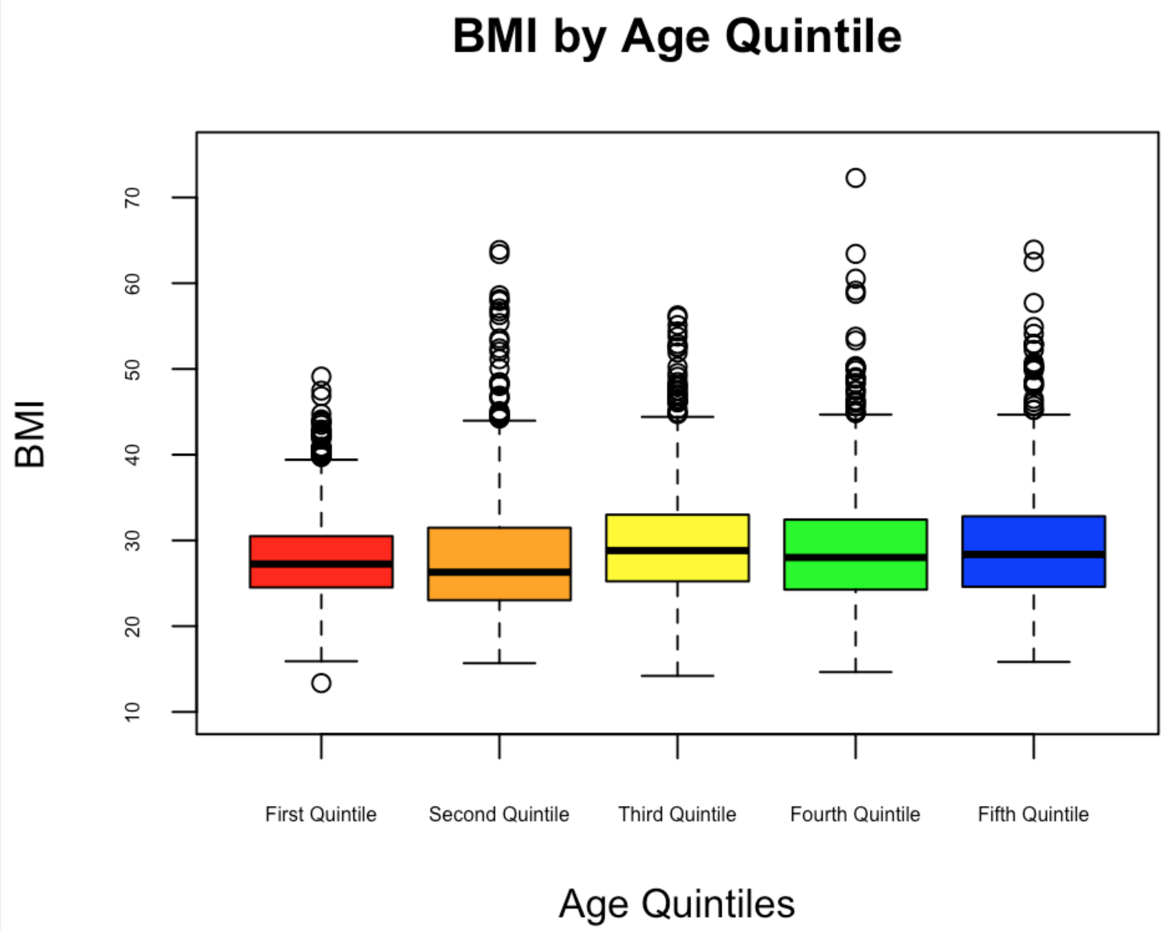


The non-transformed BMI distribution is slightly normal, but skewed to the right. The data might be more normal if log-transformed.



As expected, the log-transformed BMI appears much more normally distributed than the raw bmi.

Problem 3.



First Quintile Mean = 27.92584
Second Quintile Mean= 29.10126
Third Quintile Mean= 29.54441
Fourth Quintile Mean=29.66864
Fifth Quintile Mean= 27.93290

We can see that, as a whole, BMI does not have a consistent relationship with age. From our boxplots, we see that the highest median BMI occurs at the third quintile age group, but then dips back down again for the highest two age quintiles. The second age quintile had the lowest overall median BMI and the lowest age quintile had the smallest range between BMI values for individuals in that quintile.

Code.

```
# Exercise 5A | EPID 674 SECTION 002 | Stephanie Mecham
>
> #Creating binary T2DM variable
>
> diabetic <- ifelse(a1c >= 6.5, 1, 0)
> diabetic_meds <- ifelse (dmmed == 1,1,0)
> T2DM <- ifelse (!is.na(diabetic) & (diabetic==1) | (!is.na(diabetic_meds) & (diabetic_meds
==1)) ,1, 0)
>
>#Problem One
> #Creating bar plots of T2DM by sex and race/ethnicity
>#T2DM by sex
> barplot(table(T2DM, gender), names= c("Male", "Female"), main= "Diabetes Status by Sex",
ylab= "Count", ylim=c(0,3500), xlab= "Diabetes Status", col=c(95:47), legend.text=TRUE,
args.legend = list(x="topleft", legend=c("Diabetic", "Non-diabetic"), bty = "n", ncol=2, cex=1.3),
las=1)
>
>#T2DM by race/ethnicity
> barplot(table(T2DM, raceeth), names.arg= c("Mexican-American", "Other Hispanic", "Non-
Hispanic White", "Non-Hispanic Black", "Other"), cex.main=1, cex.names=0.4, main= "Diabetes
Status by Race/Ethnicity", ylab= "Count", ylim=c(0,3500), xlab= "Diabetes Status",
col=c("darkblue", "red"), legend.text=TRUE, args.legend = list(x="topleft", legend=c("Diabetic",
"Non-diabetic"), bty = "n", ncol=5, cex=.5), las=1)
>
>
>#Problem 2
>#Assessing Normality of BMI Distribution
> hist(bmi, main= "BMI Distribution in BPA Dataset", ylab= "Count", ylim=c(0,3500), xlab=
"BMI", xlim=c(0, 65), col=c("purple"), cex=1.3)
>
>#Assessing Normality of Log-transformed BMI Distribution
>bmi_log <- log(bmi)
>hist(bmi_log, main= "Log-transformed BMI Distribution in BPA Dataset", ylab= "Count",
ylim=c(0,2000), xlab= "Log-transformed BMI", xlim=c(0, 6), col=c("purple"), cex=1.3)
>
>
>#Problem 3
>#Creating quintiles of age variable
> quantile(age, probs = seq(0, 1, 0.20))
  0%  20%  40%  60%  80% 100%
  20   31   42   54   68   85
>
```

```
>#Creating boxplots of BMI distribution by age quintile
> bpa$quant[age <=31] <- "First Quintile"
> bpa$quant[age >31 & age <= 42] <- "Second Quintile"
> bpa$quant[age >42 & age <= 54] <- "Third Quintile"
> bpa$quant[age >54 & age <=68] <- "Fourth Quintile"
> bpa$quant[age>68] <- "Fifth Quintile"
> bpa$quant<-factor(bpa$quant, levels=c("First Quintile", "Second Quintile", "Third Quintile",
"Fourth Quintile", "Fifth Quintile")) #Setting order of quintiles on x-axis
>
>boxplot(bmi~bpa$quant, col=c("red", "orange", "yellow", "green", "blue"), ylab="BMI",
hlimcex.lab=.5, xlab="Age Quintiles", main="BMI by Age Quintile",ylim=c(10,75), cex.lab=1,
cex.axis=0.5, names=c("First Quintile", "Second Quintile", "Third Quintile", "Fourth Quintile",
"Fifth Quintile"))
>
>#Checking distribution
>aggregate(bmi, by=list(bpa$quant), mean, na.rm=T)
         Group.1 mean.bmi
1  First Quintile 27.92584
2 Second Quintile 29.10126
3  Third Quintile 29.54441
4 Fourth Quintile 29.66864
5  Fifth Quintile 27.93290
>
>
>#End of code
```