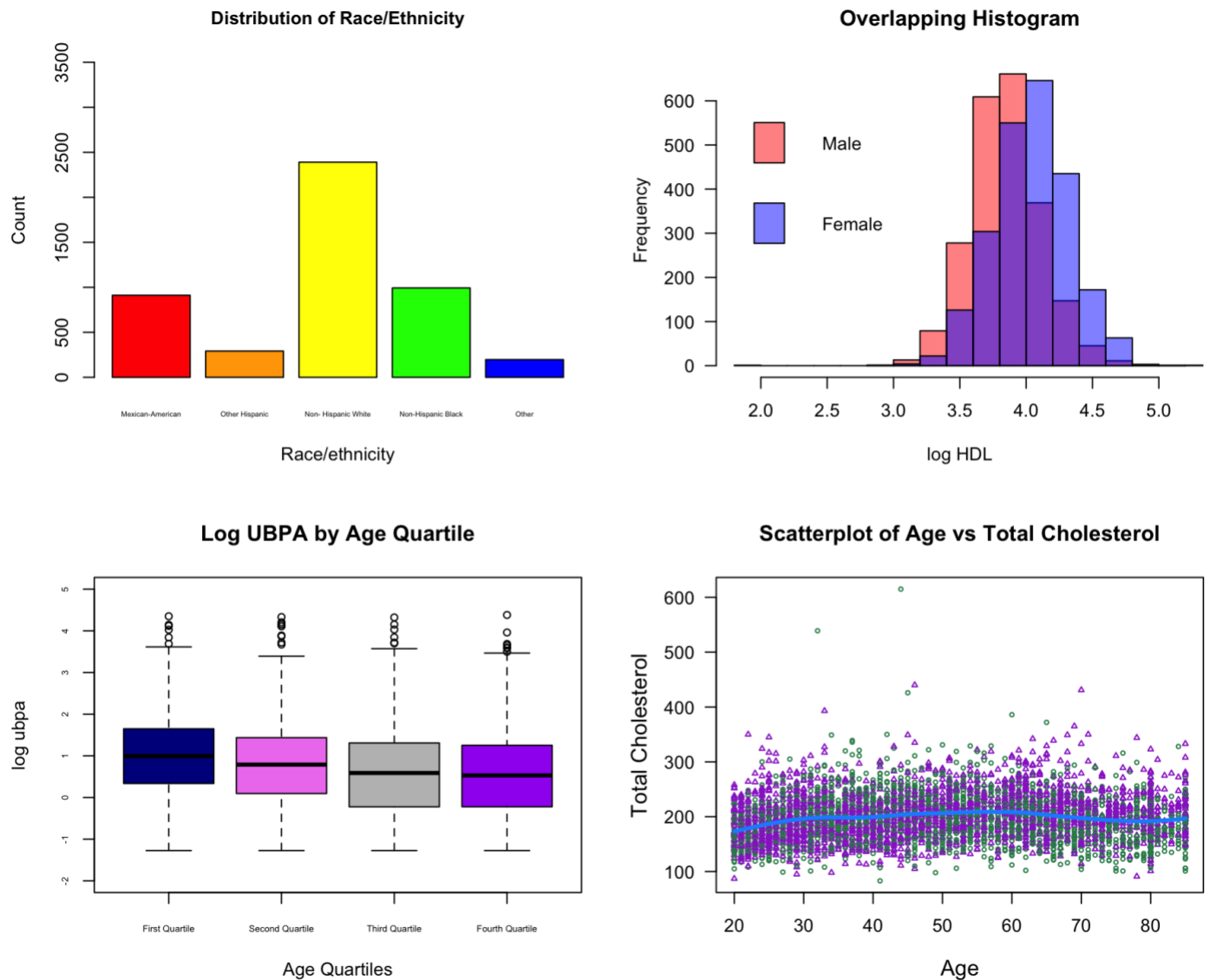


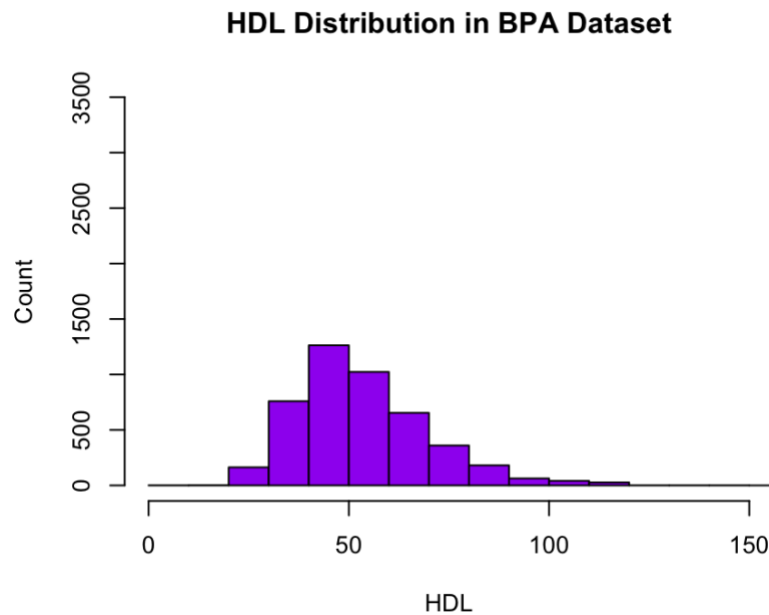
Exercise 5B:

Problem 1. Create a multi-frame plot and include 1) a bar chart of the race/ethnicity distribution; 2) a histogram of log-transformed HDL with males and females plotted separately; 3) boxplot of log-transformed urinary BPA by age quartiles; and 4) scatterplot with smoothing trend lines between age (in x-axis) and total cholesterol (in y-axis) by sex.

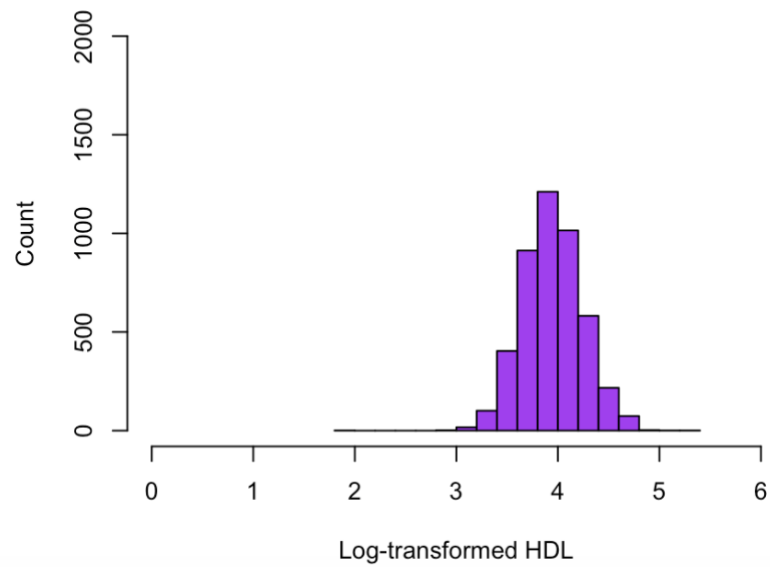


Problem 2. You want to describe the association between HDL (dependent variable, continuous) and urinary BPA (independent variable, continuous). First determine whether log-transformation for HDL is required. Create a scatterplot for HDL and urinary BPA. Add a least square line. What is the relationship between HDL and urinary BPA?

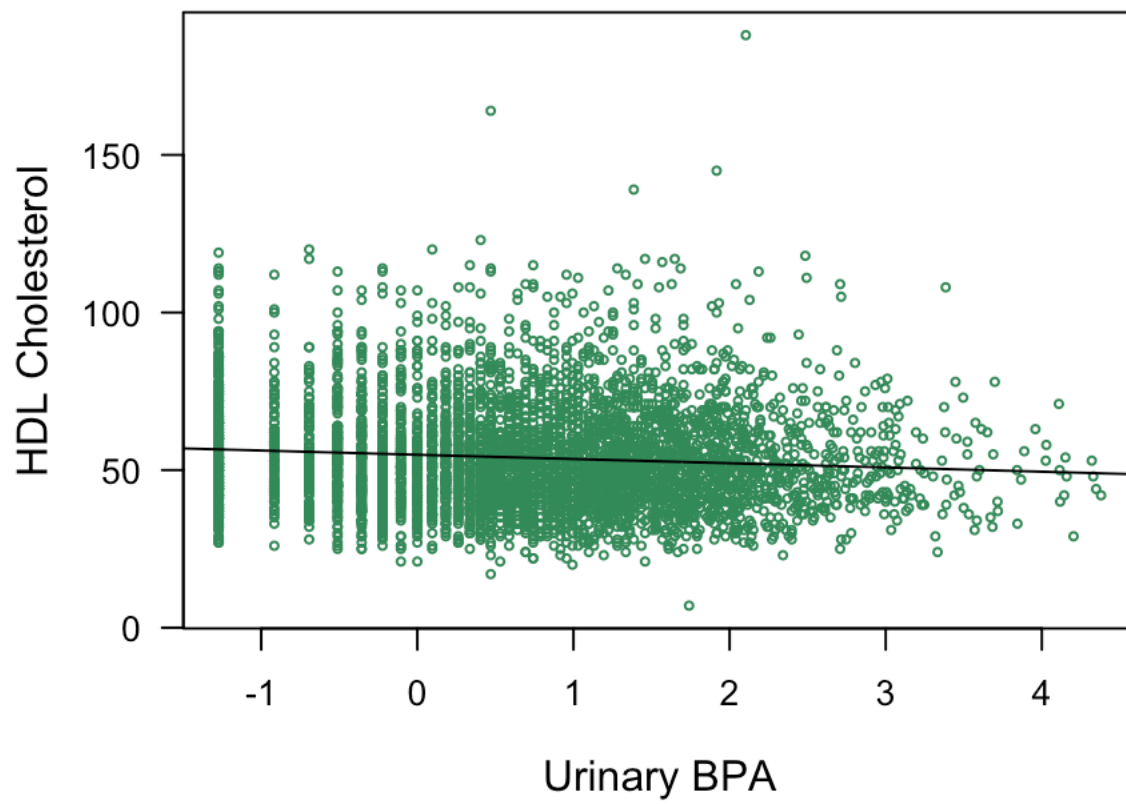
The non-transformed HDL distribution is skewed slightly right, and the distribution does appear more normal when log-transformed, but I believe the non-transformed HDL is still adequately distributed to use as is. There appears to be a slight negative linear relationship between log-transformed urinary BPA and HDL cholesterol levels.



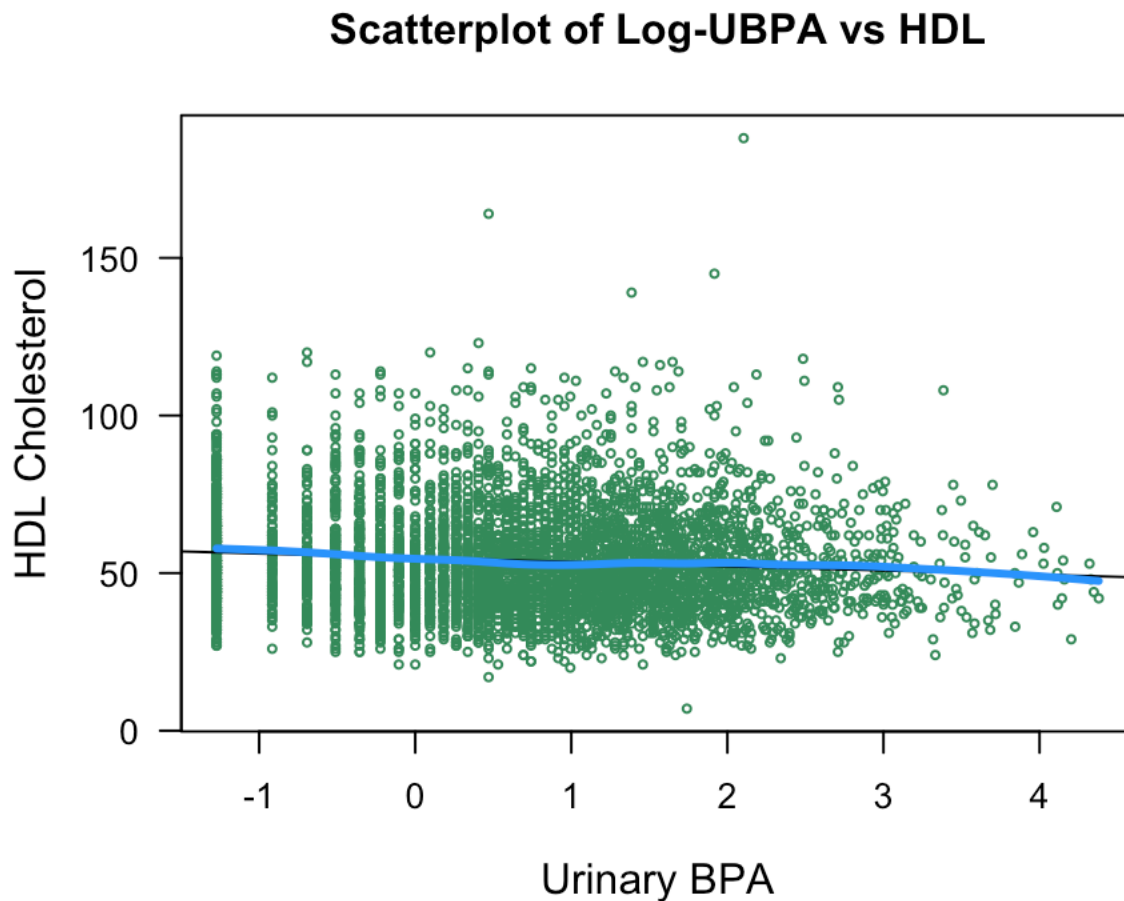
Log-transformed HDL Distribution in BPA Dataset



Scatterplot of Log-UBPA vs HDL



Problem 3. Is the association between HDL and urinary BPA linear? Check this out using a smoothing plot (lowess) from the ggplot2 package or smooth.spline() in base graphics.



It appears to be vaguely linear, but in reality the smoothing line dips down and back up again and is not entirely linear.

Code:

```
># Exercise 5B | EPID 674 Section 2 | Stephanie Mecham
>
>#Loading and attaching dataset
>
> load("/Users/smecham/Desktop/EPID 674/Datafiles & Dictionaries/bpa.rda")
> attach(bpa)
>
>#Problem One
>
> par(mfrow=c(2,2))
>
```

```

>#Bar chart
> barplot(table(raceeth), names.arg= c("Mexican-American", "Other Hispanic", "Non-
Hispanic White", "Non-Hispanic Black", "Other"), cex.main=1, cex.names=0.4, main=
"Distribution of Race/Ethnicity", ylab= "Count", ylim=c(0,3500), xlab= "Race/ethnicity",
col=c("red", "orange", "yellow", "green", "blue"))
>
>#Histogram
> log_hdl=log(hdl)
> hist(log_hdl[gender==1], col=rgb(1,0,0,0.5), main="Overlapping Histogram", xlab="log
HDL", las=1)
> hist(log_hdl[gender==2], col=rgb(0,0,1,0.5), add=T)
legend("topleft", legend=c("Male", "Female"), fill=c(rgb(1,0,0,0.5), rgb(0,0,1,0.5)), bty = "n" )
>
>#Boxplot
> log_ubpa = log(ubpa)
> quantile(age, probs=seq(0,1,0.25))
 0% 25% 50% 75% 100%
 20 33 48 65 85
> bpa$quant[age <= 33] <- "First Quartile"
> bpa$quant[age <= 33] <- "First Quartile"
> bpa$quant[age > 33 & age <= 48] <- "Second Quartile"
> bpa$quant[age > 48 & age <= 65] <- "Third Quartile"
> bpa$quant[age > 65] <- "Fourth Quartile"
> bpa$quant<-factor(bpa$quant, levels=c("First Quartile", "Second Quartile", "Third Quartile",
"Fourth Quartile")) #Setting order of quartiles on x-axis
> boxplot(log_ubpa~bpa$quant, col=c("darkblue", "violet", "gray", "purple"), ylab="log
ubpa", hlimcex.lab=.5, xlab="Age Quartiles", main="Log UBPA by Age Quartile",ylim=c(-2,5),
cex.lab=1, cex.axis=0.5, names=c("First Quartile", "Second Quartile", "Third Quartile",
"Fourth Quartile"))
>
>#Scatterplot
> plot(age, chol,xlab="Age",ylab="Total Cholesterol", pch=gender,
col=c("seagreen","darkorchid")[gender], cex=0.5, las=1, cex.lab=1.2)
> title("Scatterplot of Age vs Total Cholesterol")
> lines(smooth.spline(age[!is.na(chol)],na.omit(chol), df=10), col = "dodgerblue", lwd=3)
>
>#Problem 2
>
>#Assessing normality of HDL distribution
> > hist(hdl, main= "HDL Distribution in BPA Dataset", ylab= "Count", ylim=c(0,3500), xlab=
"HDL", xlim=c(0, 150), col=c("purple"), cex=1.3)
>
>#Assessing normality of log-transformed HDL distribution
> hdl_log <- log(hdl)

```

```
> hist(hdl_log, main= "Log-transformed HDL Distribution in BPA Dataset", ylab= "Count",
ylim=c(0,2000), xlab= "Log-transformed BMI", xlim=c(0, 15), col=c("purple"), cex=1.3)
>
>#Creating a Scatterplot
> >plot(log_ubpa, hdl,xlab="Urinary BPA",ylab="HDL Cholesterol", col=c("seagreen"),
cex=0.5, las=1, cex.lab=1.2)
>
>#Adding Least Squares Line
> fit <- lm(hdl~log_ubpa)
> abline(fit)
> title("Scatterplot of Log-UBPA vs HDL")
>
>
>#Problem 3
> lines(smooth.spline(log_ubpa[!is.na(hdl)],na.omit(hdl), df=10), col = "dodgerblue", lwd=3)
>
>#End of code
```