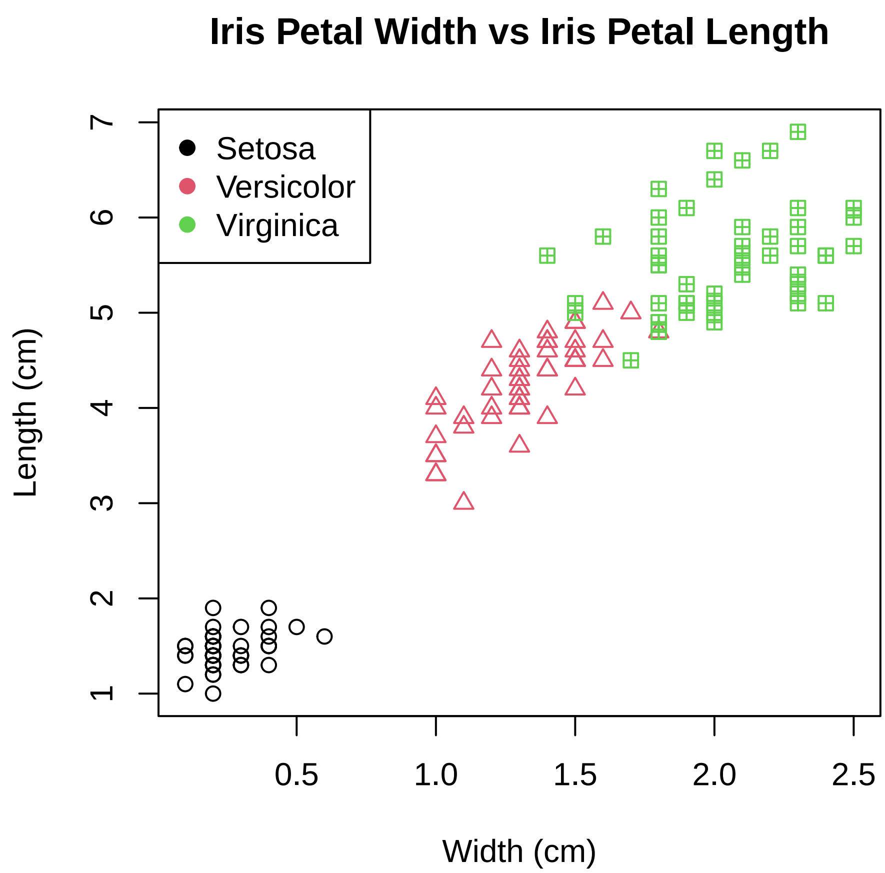
1 Plotting Iris Dataset

The code and output for this section is provided below:

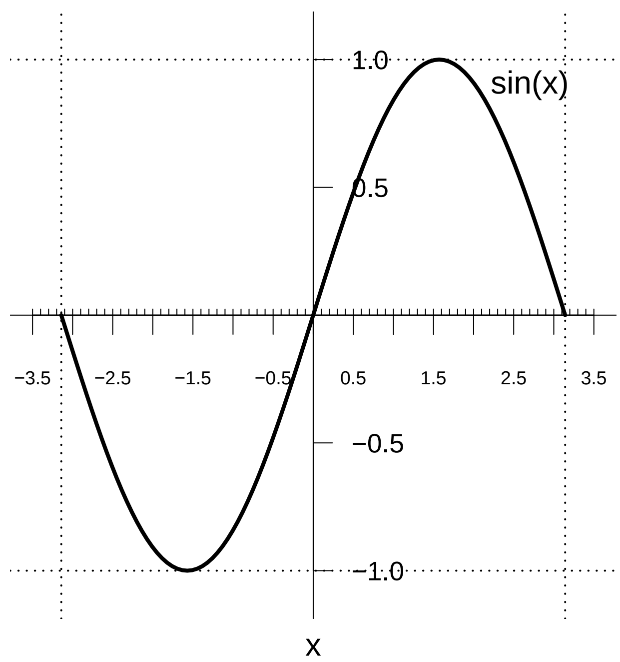
```
plot(iris$Petal.Width, iris$Petal.Length,
     xlab = "Width (cm)",
     ylab = "Length (cm)",
     pch = c(Sertosa = 1, Versicolor = 2, Virginica = 12)[iris$Species],
     col = iris$Species)

title("Iris Petal Width vs Iris Petal Length")
legend("topleft",
legend=c("Setosa", "Versicolor", "Virginica"), col=c(1,2,3), pch=19)
```



2 Plot of Sin(x)

The code and output for this section is provided below:

```
par(pty = "s") # Making sure its square
plot(sin, from = -pi, to = pi, xlim = c(-3.5,3.5), ylim = c(-1.1, 1.1),
     yaxt = "n", xaxt = "n", bty = "n", xlab = "", ylab = "", lwd = 2)

abline(h = 0, v = 0, lwd = 0.5)
abline(h = 1, v = pi, lty = 3)
abline(h = -1, v = -pi, lty = 3)

axis(side = 1, at = seq(-3.5, 3.5, 0.5), pos = 0, lwd = 0.5,
     lwd.ticks = 0.5, cex.axis = 0.6, line = -0.9)
axis(side = 1, at = seq(-3.5, 3.5, 0.1), labels = FALSE, pos = 0,
     lwd = 0.5, lwd.ticks = 0.5, tck = 0.01)
axis(side = 4, at = c(-1, -0.5, 0.5, 1), pos = 0, lwd = 0.5,
     lwd.ticks = 0.5, las = 2, cex.axis = 0.8)

text(2.7,0.9,labels="sin(x)")
text(0,-1.3,labels="x", xpd=TRUE)
```



3 CDF of X

a. Express CDF of Y as CDF of X

CDF of Y: $F_Y(y) = P(Y < y) = P(\sigma X - \mu < y) = P(X < \dfrac{y+\mu}{\sigma})$

$$F_Y(y) = F_X(\frac{y+\mu}{\sigma})$$

b. Express Quantile of Y as Quantile of X

$$Q_Y(p) = F_Y^{-1}(p) \implies F_Y(Q_Y(p)) = p,$$

$$\therefore F_X\left(\frac{Q_Y(p)+\mu}{\sigma}\right) = p.$$

$$\frac{Q_Y(p)+\mu}{\sigma} = F_X^{-1}(p), \text{ since F(x) is invertible}$$

$$Q_Y(p) = \sigma F_X^{-1}(p) - \mu$$

$$Q_Y(p) = \sigma Q_X(p) - \mu$$

4 Kernel Density Estimator

a. Show that it is non-negative

b, n $> 0 \therefore \frac{1}{bn} > 0$

$K(u) > 0$ for all u, by definition:

$$\begin{aligned} K(u) &> 0 \\ \implies \sum_i K(u) &> 0 \\ \implies \frac{1}{nb}\sum_i K(u) &> 0 \\ \therefore \frac{1}{nb}\sum_i K\left(\frac{y_i - x}{b}\right) &> 0 \\ \therefore \hat{f}(x) &> 0 \end{aligned}$$

b. Show that it integrates to 1

$$\begin{aligned} \int \hat{f}(x) &= \int \frac{1}{nb}\sum_i K\left(\frac{y_i - x}{b}\right) dx \\ &= \frac{1}{nb}\sum_i \int K(u)\, dx, \text{ for } u = \frac{y_i - x}{b} \\ &= \frac{1}{nb}\sum_i \int K(u)\, b\, du, \text{ for } dx = b\, du \\ &= \frac{1}{n}\sum_i 1, \text{by definition of K} \\ &= \frac{n}{n} = 1 \end{aligned}$$

5 Silverman's Rule

For this question, I decided to use the Iris Dataset in R, but using Sepal Length instead of Petal.

a. Compute Bandwidth using density() function

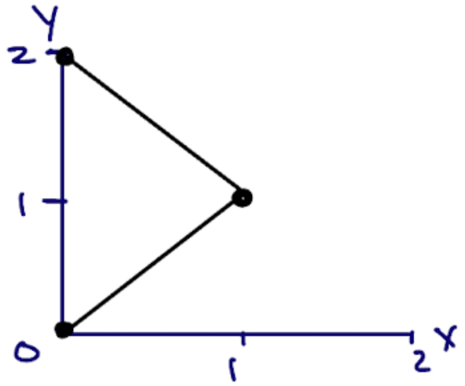
```
density(iris[["Sepal.Length"]], bw="nrd0")

##
## Call:
##  density.default(x = iris[["Sepal.Length"]], bw = "nrd0")
##
## Data: iris[["Sepal.Length"]] (150 obs.); Bandwidth 'bw' = 0.2736
##
##        x               y
##  Min.   :3.479   Min.   :0.0001495
##  1st Qu.:4.790   1st Qu.:0.0341599
##  Median :6.100   Median :0.1534105
##  Mean   :6.100   Mean   :0.1905934
##  3rd Qu.:7.410   3rd Qu.:0.3792237
##  Max.   :8.721   Max.   :0.3968365

density(iris[["Sepal.Length"]], bw="nrd")

##
## Call:
##  density.default(x = iris[["Sepal.Length"]], bw = "nrd")
##
## Data: iris[["Sepal.Length"]] (150 obs.); Bandwidth 'bw' = 0.3222
##
##        x               y
##  Min.   :3.333   Min.   :0.0001447
##  1st Qu.:4.717   1st Qu.:0.0250724
##  Median :6.100   Median :0.1382423
##  Mean   :6.100   Mean   :0.1805410
##  3rd Qu.:7.483   3rd Qu.:0.3605404
##  Max.   :8.867   Max.   :0.3935263
```

b. Compute Bandwidth using Formula

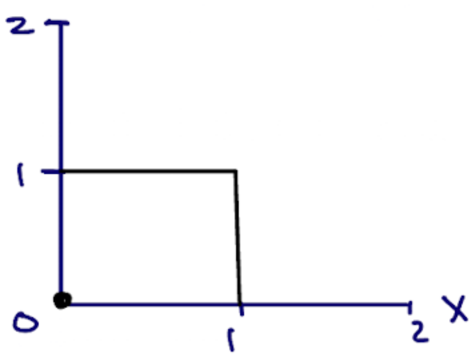
```
data <- iris[["Sepal.Length"]]
(0.9 * min(sd(data), IQR(data)/1.34))*(length(data)^(-1/5))

## [1] 0.2735831

(1.06 * min(sd(data), IQR(data)/1.34))*(length(data)^(-1/5))

## [1] 0.3222201
```

As seen above, the formula derived in class yields the same bandwidth as the density() functions.

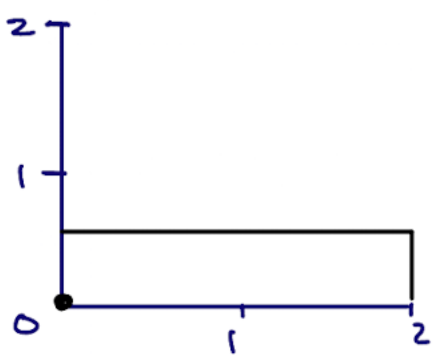
6 Triangle Distribution



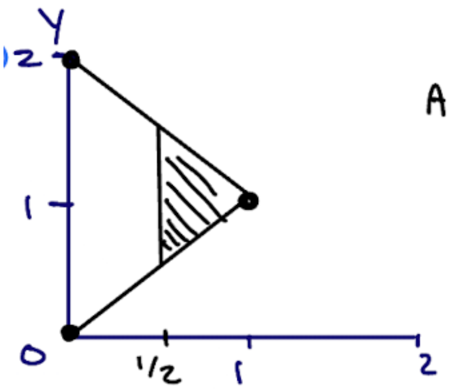
a. Marginal Density of X



b. Marginal Density of Y

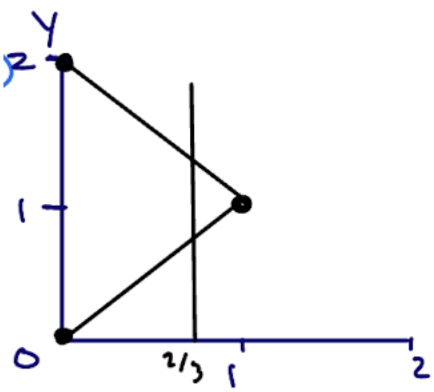


c. First Probability



We want $P[X + Y \leq 1]$, which is demonstrated by the shaded region above on the joint density graph. We know that area under density represents probability, and also that probability of a triangle is $\frac{1}{2}\, b\, h$.
$\therefore P[X + Y \leq 1] = A = \frac{1}{2}bh = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2} + \frac{1}{2}\right) = \frac{1}{4}$

d. Second Probability



We want to know $P[Y > 1/3 | X = 2/3]$. As seen above, when X = 2/3, Y is always ¿ 1/3.
$\therefore P[Y > 1/3 | X = 2/3] = 1$