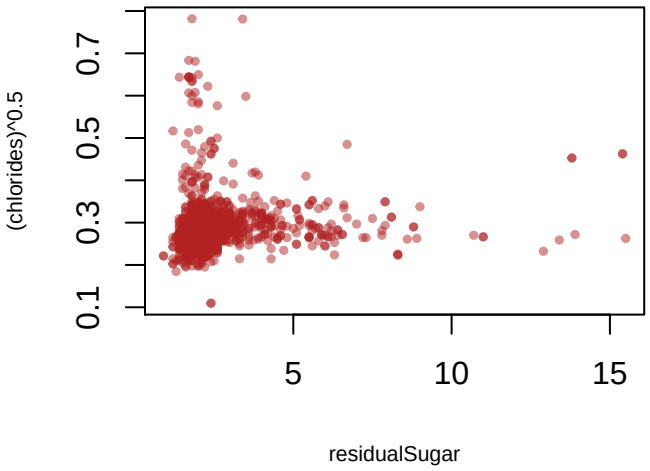
A2Q2

Dataset: Wine_quality (Click here)

(a) (2 *marks*) Consider the scatterplot below of $(chlorides)^{1.5}$ versus $residualSugar$ from wine quality data set:

Which of the following transformations might straighten this scatterplot?



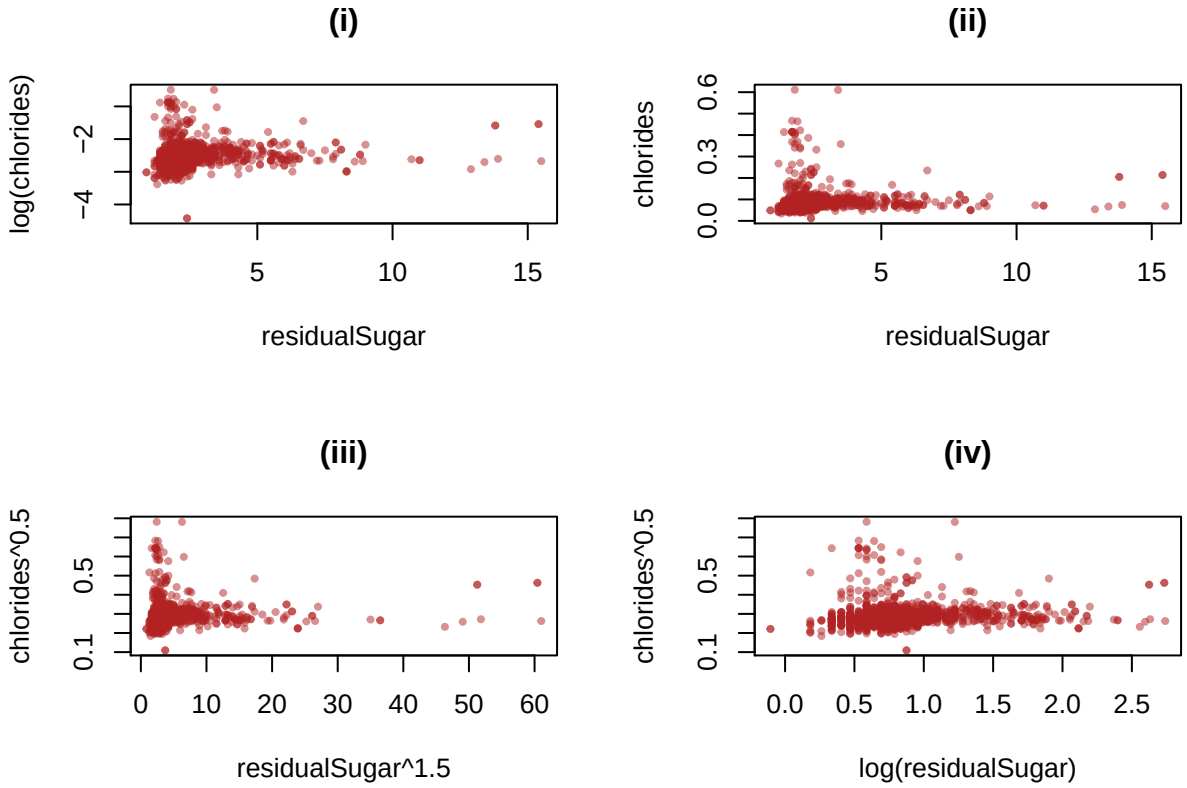
(i) $log(chlorides)$ vs $residualSugar$

(ii) $chlorides$ vs $residualSugar$

(iii) $(chlorides)^{0.5}$ vs $(residualSugar)^{1.5}$

(iv) $(chlorides)^{0.5}$ vs $log(residualSugar)$

Answer: (i) (iv)

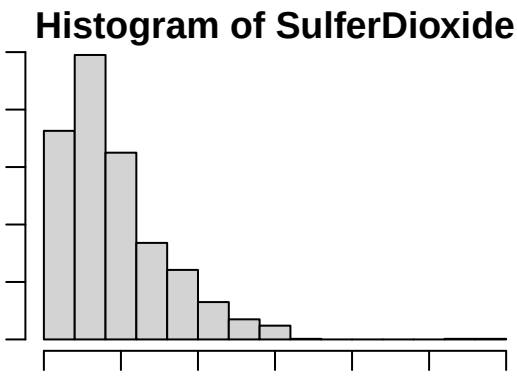


Explanation:

From the original graph, we can see that the curvature corresponds to the third quadrant. This suggests that the ladder move should be "Down on X; Down on Y". For (i), α of $residualSugar$ remains the same while α of $chlorides$ decreases. For (ii), α of $residualSugar$ remains the same, but α of $chlorides$ increases. For (iii), α of $residualSugar$ increases, but α of $chlorides$ remains the same. For (iv), α of $residualSugar$ decreases, but α of $chlorides$ remains the same. Therefore, (i) and (iv) are true.

(b) (2 *marks*) Consider the histogram below of $SulferDioxide$:

Which of the following transformations might makethis histogram more symmetric?



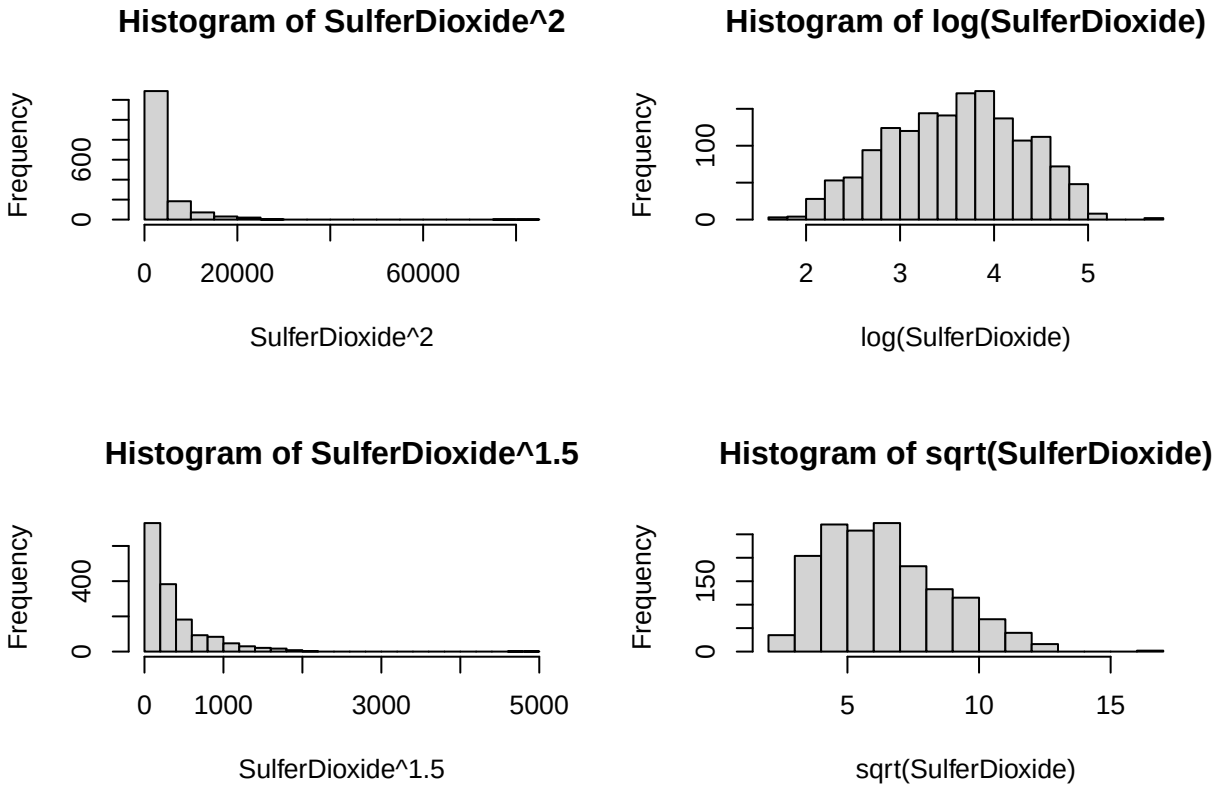
(i) $SulferDioxide^2$

(ii) $log(SulferDioxide)$

(iii) $SulferDioxide^{1.5}$

(iv) $\sqrt{SulferDioxide}$

Answer: (ii) (iv)



Explanation:

From the original graph, we see that the "bump" is on lower values. This suggests us to move the power "lower" on the ladder. Since we have α equals to 1. To make the graph symmetric, we want α to be smaller than. In these 4 options, (ii) has α=0 and (iv) has 1/2. Therefore, (ii) and (iv) are correct. From graphs, we can see that the right tails of distributions in (ii) and (iv) are reduced and the bumps move to the right. For the other two graphs, the right tails become longer, distributions are more asymmetric.

Appendix

```
wine <- read.csv("winequality-red.csv", header=TRUE)
residualSugar <- wine$residual.sugar
chlorides <- wine$chlorides
par(mar=c(5,5,2,2))
plot(residualSugar, (chlorides)^0.5, pch = 19,
     col=adjustcolor("firebrick", alpha = 0.5 ), cex = 0.5, cex.lab = 0.7)
```

```
par(mfrow=c(2,2))
plot(residualSugar, log(chlorides), pch = 19,
     col=adjustcolor("firebrick", alpha = 0.5 ), cex=0.5, main="(i)")
plot(residualSugar, chlorides, pch = 19,
     col=adjustcolor("firebrick", alpha = 0.5 ), cex=0.5, main="(ii)")
plot(residualSugar^1.5, chlorides^0.5, pch = 19,
     col=adjustcolor("firebrick", alpha = 0.5 ), cex=0.5, main="(iii)")
plot(log(residualSugar), chlorides^0.5, pch = 19,
     col=adjustcolor("firebrick", alpha = 0.5 ), cex=0.5, main="(iv)")
```

```
par(mar=c(1,1,1,1))
SulferDioxide <- wine$total.sulfur.dioxide
hist(SulferDioxide, breaks = 20)


par(mfrow=c(2,2))
```

```
hist(SulferDioxide^2, breaks = 20)
hist(log(SulferDioxide), breaks = 20)
hist(SulferDioxide^1.5, breaks = 20)
hist(sqrt(SulferDioxide), breaks = 20)
```