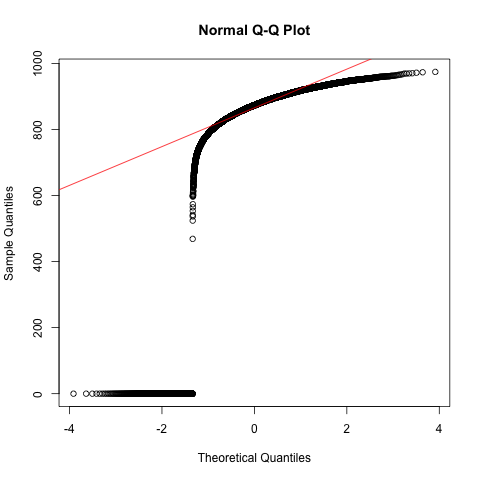
# Assignment 10

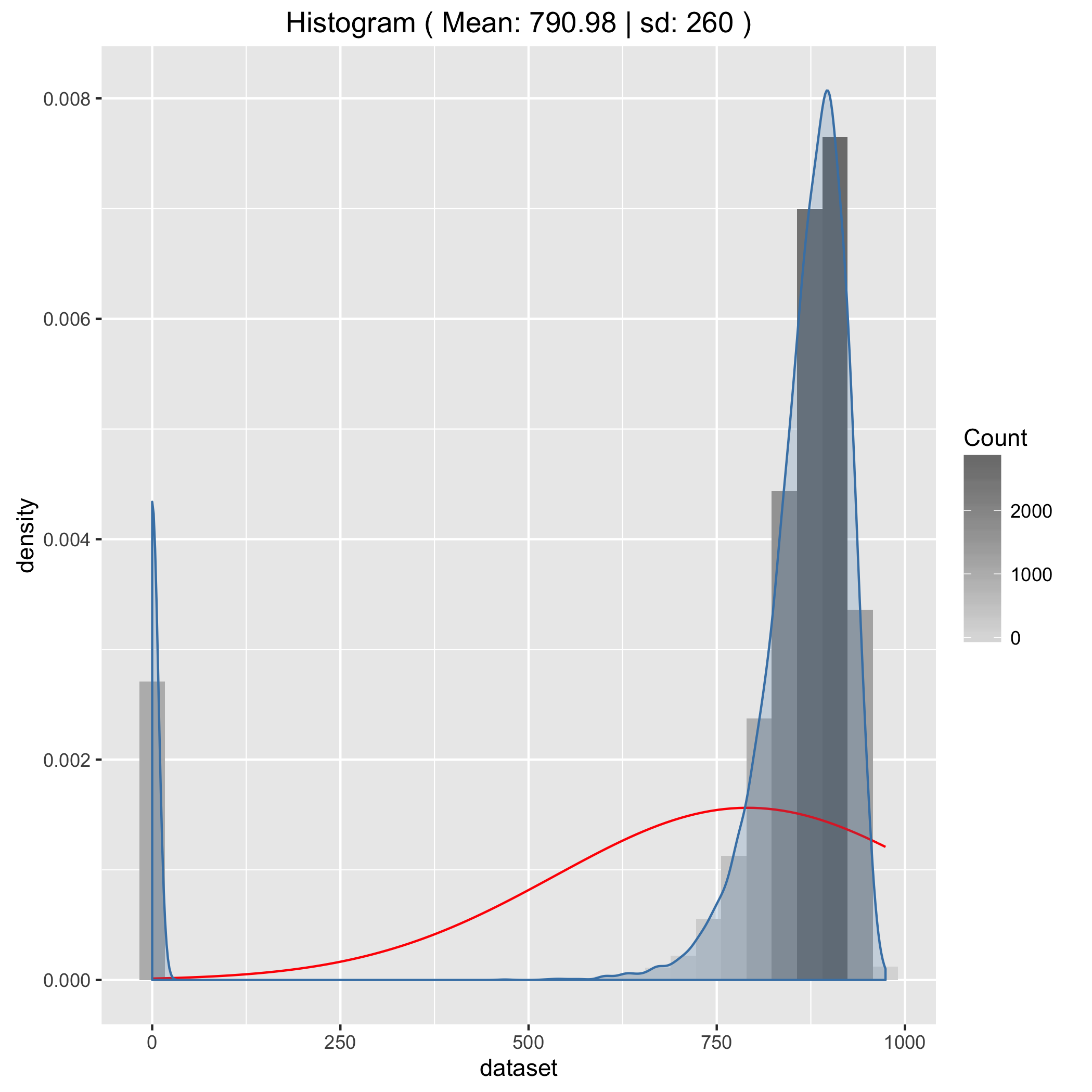
### Exercise 1.8.1

##### Question 1

1. A different cell population is simulated in dataset cell\_population\_b also contained in the file cell\_fluorescence.RData. Verify that this population does not follow a normal distribution.

* # test can only take a maximum of 5000 datapoints, so take sample   
  shapiro.test(sample(cell\_population\_b,5000))   
  # Shapiro-Wilk normality test  
  # data: sample(cell\_population\_b, 5000)  
  # W = 0.49479, p-value < 2.2e-16  
    
  png(filename="Assignment/qqnorm.png")  
  qqnorm(cell\_population\_b)  
  qqline(cell\_population\_b, col='red')   
  dev.off()  
    
  dist\_mean <- mean(cell\_population\_b)  
  dist\_sd <- sd(cell\_population\_b)  
    
  gg <- ggplot(as.data.frame(cell\_population\_b), aes(cell\_population\_b))  
  gg <- gg + geom\_histogram(aes(y=..density.., fill=..count..))  
  gg <- gg + scale\_fill\_gradient("Count", low="#DCDCDC", high="#7C7C7C")  
  gg <- gg + stat\_function(fun=dnorm, color="red",  
   args=list(mean=dist\_mean,sd=dist\_sd))  
  # Adds a density plot on top  
  gg <- gg + geom\_density(alpha = 0.2, fill="steelblue", colour="steelblue")   
  gg <- gg + ggtitle(paste("Histogram", "( Mean:", round(dist\_mean,2),   
   "sd:", signif(dist\_sd,2), ")"))  
    
  gg  
  ggsave('Assignment/histogram.png')





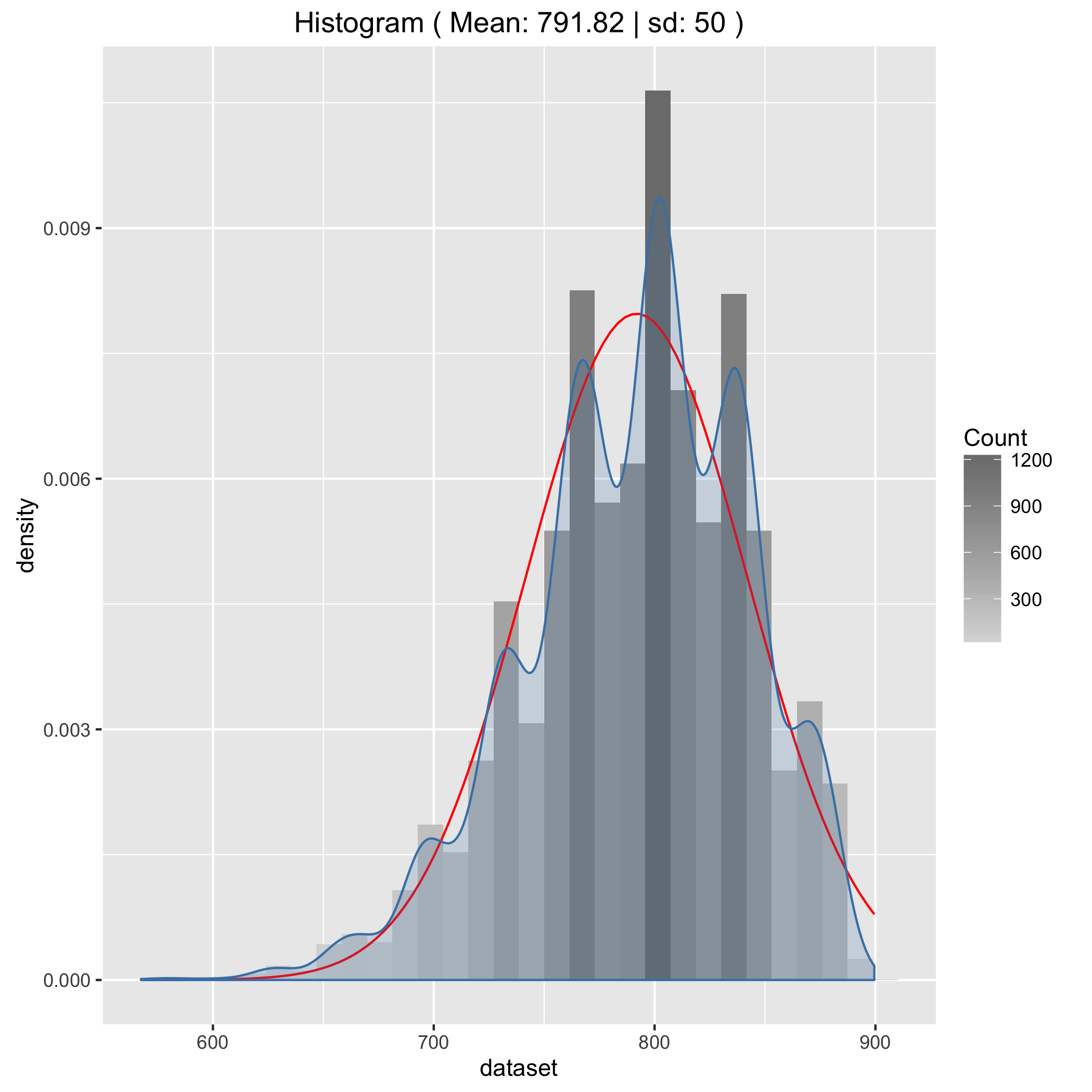
##### Question 2

1. Describe the fluorescence of this cell population, with reference to its distribution shape.

* As shown by the Q-Q plot above together with the histogram below, we can see that the distribution is definitely not normal. Rather, the histogram shows that the distribution of the flourence is skewed to the right with a long tail on the left.

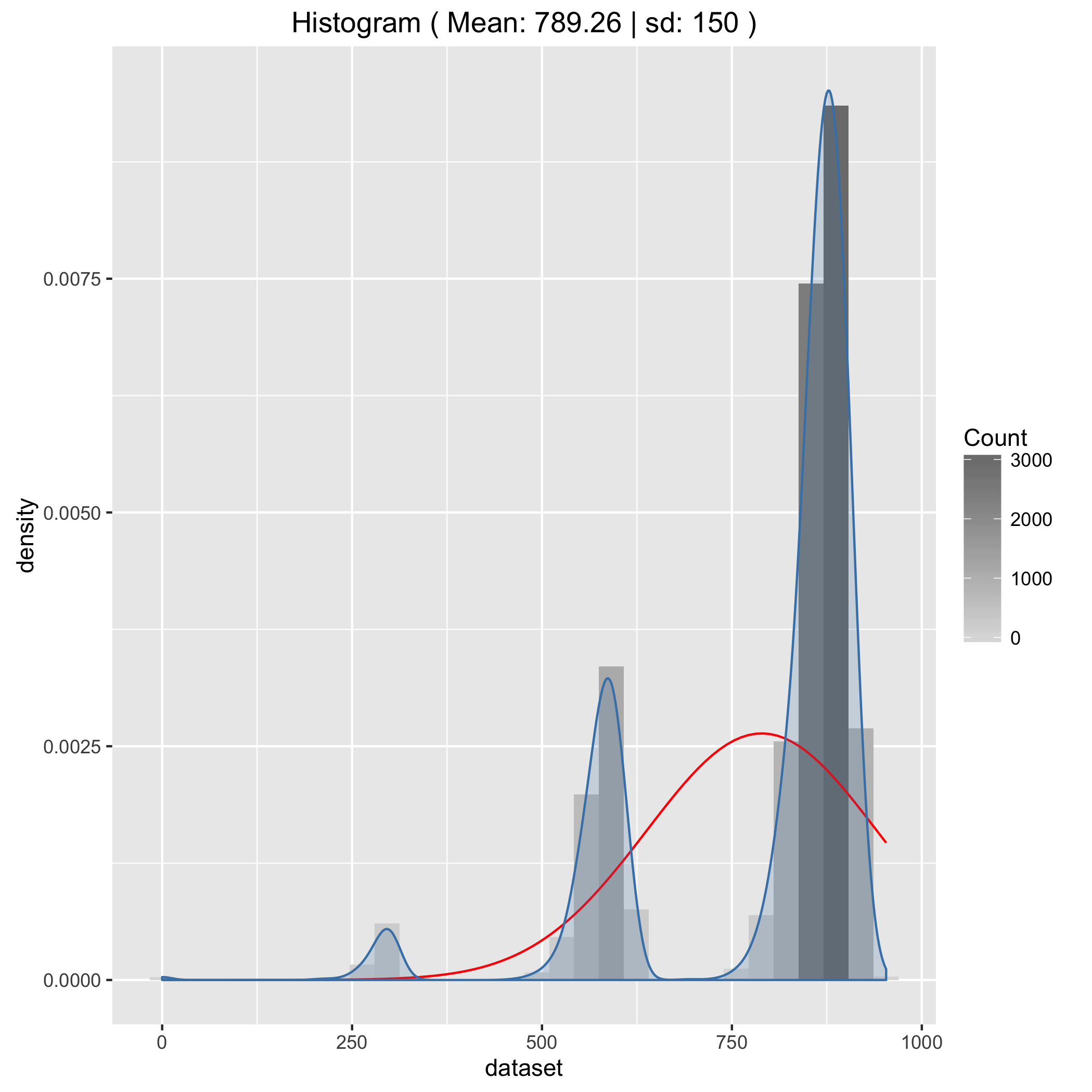
##### Question 3

1. To investigate the prediction of the Central Limit Theorem write a function repeat\_measurements which repeats many measurements on an input population dataset, then plots the distribution of sample means.

* repeat\_measurements <- function (dataset, sample\_size=25,  
   n\_repeats=10000) {  
   mean\_distribution <- sapply(1:n\_repeats, function(r)  
   mean(sample(dataset,sample\_size))  
   )  
   dist\_mean <- mean(mean\_distribution)  
   dist\_sd <- sd(mean\_distribution)  
   data <- as.data.frame(mean\_distribution)  
   gg <- ggplot(data, aes(x = mean\_distribution))   
   gg <- gg + geom\_histogram(aes(y=..density.., fill=..count..))  
   gg <- gg + scale\_fill\_gradient("Count", low="#DCDCDC", high="#7C7C7C")  
   gg <- gg + stat\_function(fun=dnorm, color="red",  
   args=list(mean=dist\_mean,sd=dist\_sd))  
   gg <- gg + geom\_density(alpha = 0.2, fill="steelblue", colour="steelblue")  
   gg <- gg + ggtitle(paste("Histogram", "( Mean:", round(dist\_mean,2),   
   "sd:", signif(dist\_sd,2), ")"))  
   return( list(graph=gg, mean=dist\_mean, sd=dist\_sd) )  
  }  
    
  repeat\_measurements(cell\_population\_b)  
  ggsave('Assignment/function\_histogram.png')
* 

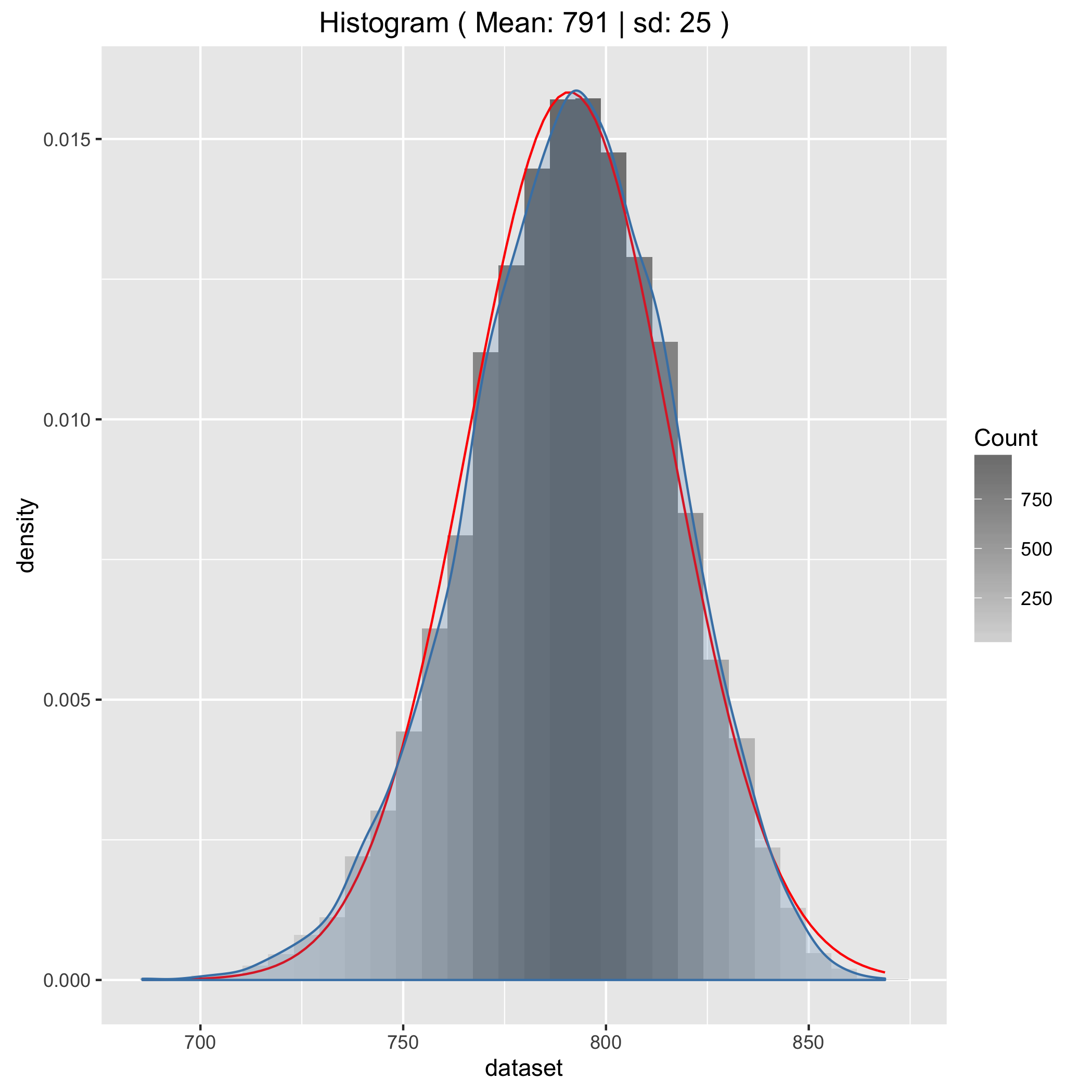
##### Question 4

1. Use this function to investigate the behaviour of the sample mean distribution for a low sample sizes of cell\_population\_b (e.g. sample\_size=3)

* repeat\_measurements(cell\_population\_b, 3, 10000)  
  ggsave('Assignment/function\_histogram\_sample\_size\_3.png')
* 

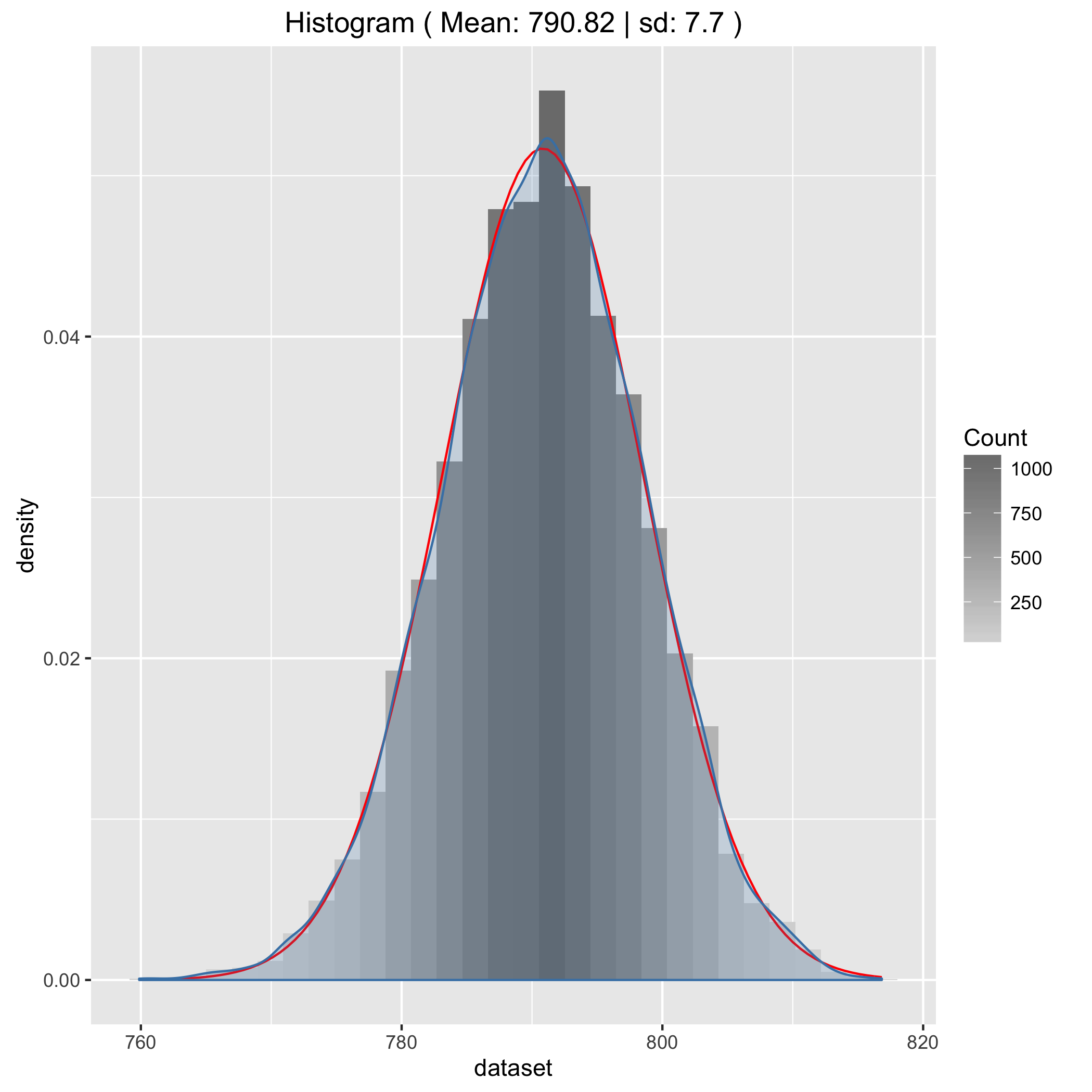
##### Question 5

1. Next identify a sample size that is large enough so that the distribution looks approximately normal. Note the mean and standard deviation of this distribution.

* repeat\_measurements(cell\_population\_b, 100, 10000)  
  ggsave('Assignment/almost\_normal\_histogram.png')
* 

##### Question 6

1. Increase the sample size by a factor of 10 and note the new values for mean and standard deviation. Is this in line with the central limit theorem?

* repeat\_measurements(cell\_population\_b, 1000, 10000)  
  ggsave('Assignment/histogram\_sample\_size\_1000.png')
* 
* Yes this is inline with the central limit theorem. A 10 fold increase in the sample size should decrease the standard deviation and the width of the distribution by a factor of roughly 3 (i.e. ).

### Exercise 2.1.9

##### Question 1

1. We will next try to determine the optimum field of view which gives least uncertainty in our final result. To aid us let's first construct a table showing our input parameters over a range of image fields sizes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| * Image Area | * Cells Imaged | * Pixels per Cell | * Amplification | * σ Pixel Noise |
| * default | * 16 | * 1024 | * 1x | * 7.5 |
| * 2x | * 32 | * 512 | * 2x | * 15 |
| * 4x | * 64 | * 256 | * 4x | * 30 |
| * 8x | * 128 | * 128 | * 8x | * 60 |
| * 16x | * 256 | * 64 | * 16x | * 120 |
| * 32x | * 512 | * 32 | * 32x | * 240 |
| * 64x | * 1024 | * 16 | * 64x | * 480 |
| * 128x | * 2048 | * 8 | * 128x | * 960 |
| * 256x | * 4096 | * 4 | * 256x | * 1920 |

##### Question 2

1. For each image field size use repeat\_measurements\_inc\_noise or your own function to calculate the standard deviation of the distribution of resulting measurements.

##### Question 3

1. Comment on the best choice of image field for this experiment.

##### Question 4

1. In a real situation what else could you do to reduce the uncertainty in the fluorescence measurements?