

Q3

```
library(fitdistrplus)
library(ggplot2)
library(reshape2)
dir = "E:\\Google Drive\\...\\Assignment 3\\"
file1 = "ONCHLOR.csv"
file2 = "Table3.csv"
dfChlor = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')
dfTable3 = read.table(file=paste(dir,file2, sep=""), header=TRUE, sep=',')
dfTable3 = dfTable3[order(dfTable3$Concentration..mg.Cl..L..), ]
dfTable3$frac = ppoints(dfTable3$Concentration..mg.Cl..L.., 0.5)
```

a)

```
fit2 = fitdist(dfTable3$Concentration..mg.Cl..L.., 'lnorm')
summary(fit2)
```

Summary of fitted model

```
Fitting of the distribution 'lnorm' by maximum likelihood
Parameters :
      estimate Std. Error
meanlog 8.3974131  0.1527407
sdlog    0.6109629  0.1080027
Loglikelihood: -149.1781  AIC: 302.3562  BIC: 303.9014
Correlation matrix:
      meanlog      sdlog
meanlog 1.000000e+00 -5.860702e-11
sdlog   -5.860702e-11 1.000000e+00
```

```
hc5_2 = quantile(fit2, probs = 0.05)
hc5_2
```

HC5

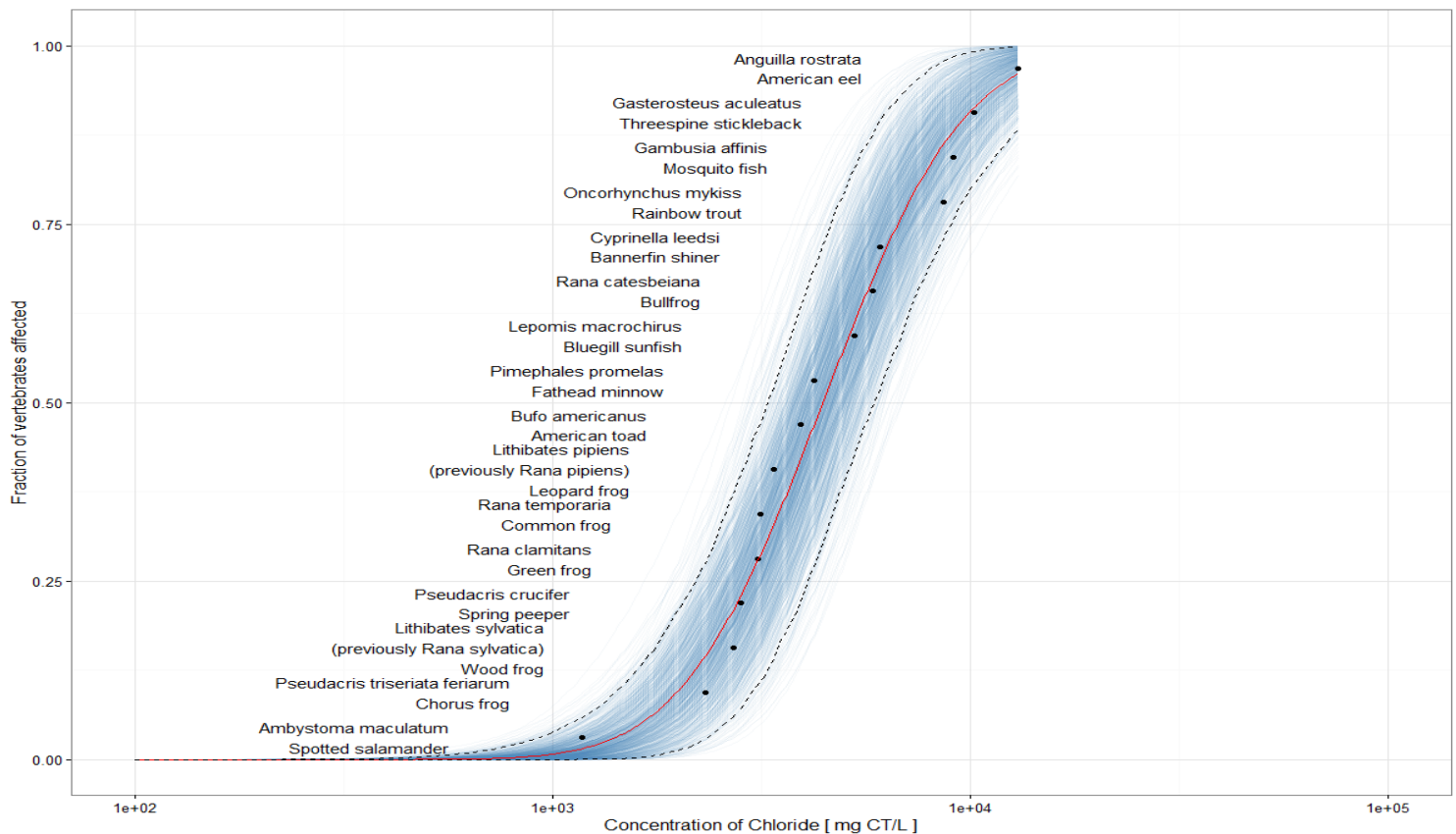
```
Estimated quantiles for each specified probability (non-censored data)
p=0.05
estimate 1623.709
```

```
fit2_bootp = bootdist(fit2, bootmethod = 'param', niter = 1000)
quantile(fit2_bootp, probs = 0.05)
```

95% parametric bootstrapped CI

```
two-sided 95 % CI of each quantile
p=0.05
2.5 % 1047.845
97.5 % 2682.789
```

- The estimated 95% parametric bootstrapped CI for is (1047.845, 2682.789)
- The estimated HC5 is 163.709
- The short term HC5 identified in the article was 640 mg CI'/L across all species and the long term was 120 mg CI/L. Using the short term number, the Northern riffleshell mussel and then Water flea fell below the HC5.



b)

```
dfChlor$frac = ppoints(dfChlor$Conc,0.5)
fit3 = fitdist(dfChlor$Conc,'lnorm')
summary(fit3)
```

Summary of fitted model

Fitting of the distribution 'lnorm' by maximum likelihood

Parameters :

	estimate	Std. Error
meanlog	3.308611	0.013431798
sdlog	1.148085	0.009497683

Loglikelihood: -35548.4 AIC: 71100.8 BIC: 71114.59

Correlation matrix:

	meanlog	sdlog
meanlog	1	0
sdlog	0	1

```
fit3_bootp = bootdist(fit3, bootmethod = 'param', niter = 1000)
quantile(fit3_bootp, probs = 0.95)
```

95% parametric bootstrapped CI

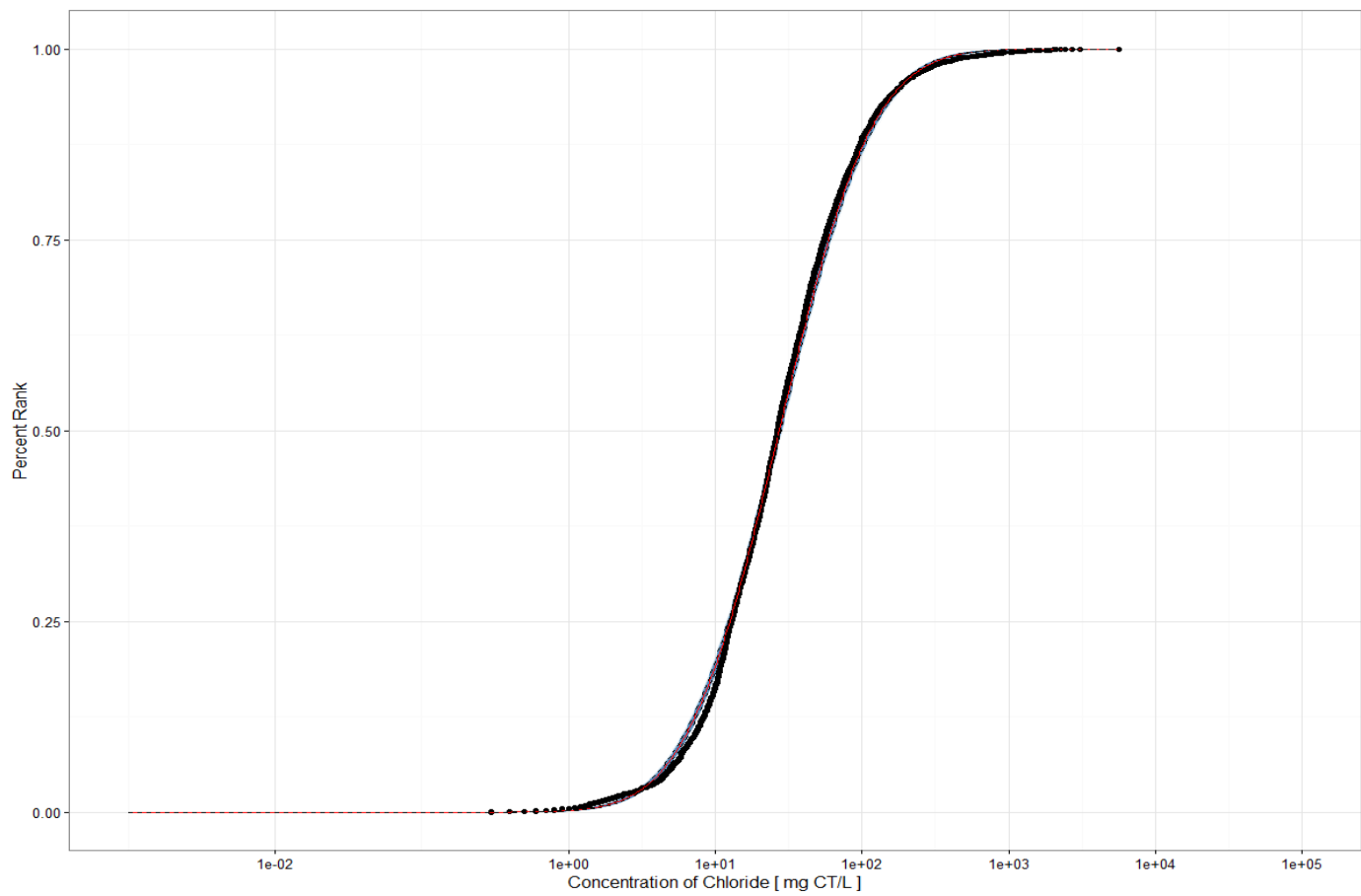
two-sided 95 % CI of each quantile

p=0.95

2.5 % 173.4731

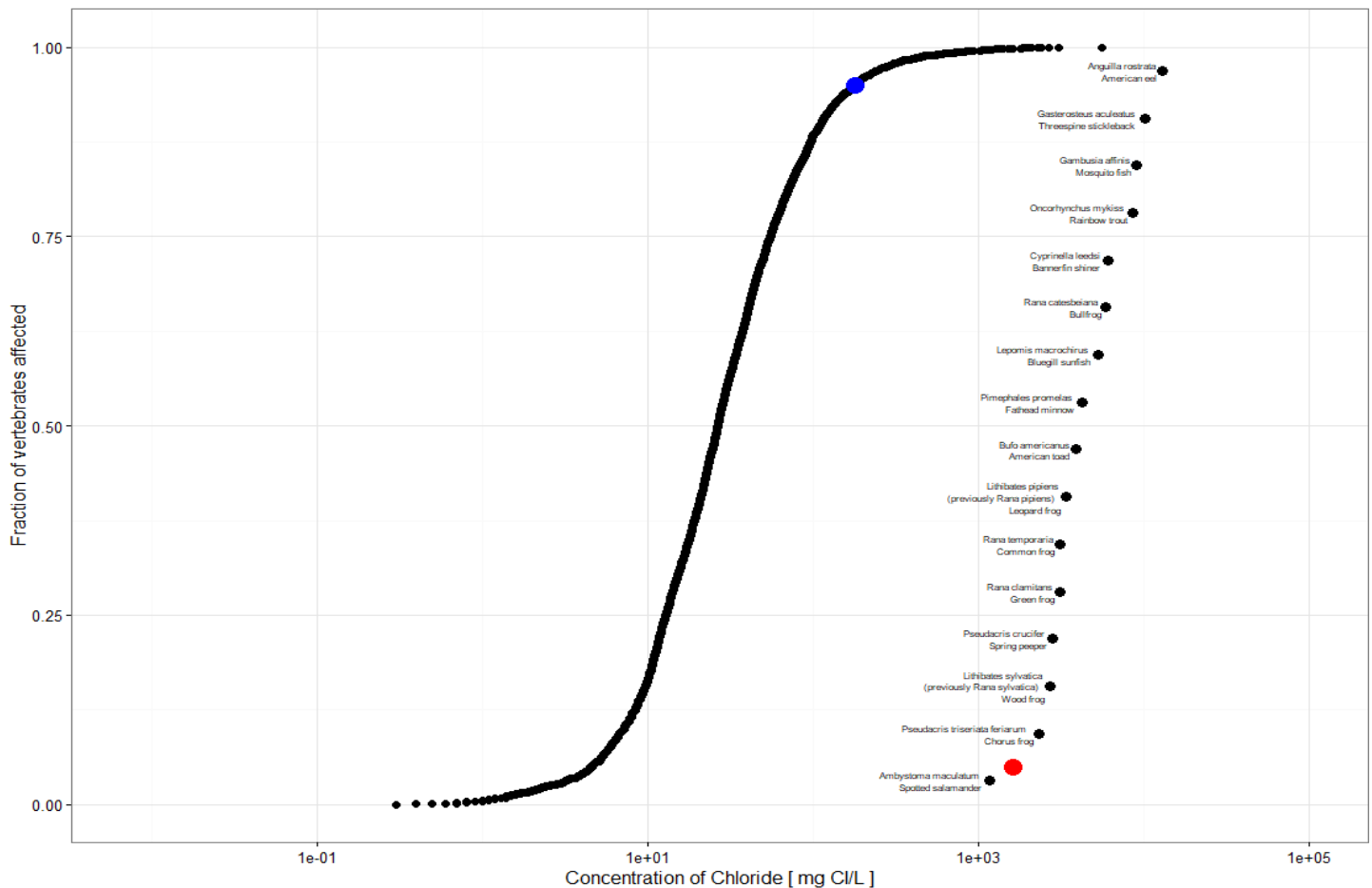
97.5 % 188.5119

- The estimated 95% parametric bootstrapped CI for is (173.4731, 188.5119)



c)

```
ggplot(data = dfTable3) +
  geom_point(aes(x = Concentration..mg.Cl..L..., y = frac), size = 3) +
  geom_point(data = dfChlor, aes(x = Conc, y = frac), size = 2) +
  geom_point(aes(x=hc5_2$quantiles[1], y = 0.05), size = 5, col = "red") +
  geom_point(aes(x=hc95_3$quantiles[1,1], y = 0.95), size = 5, col = "blue") +
  geom_text(aes(x = Concentration..mg.Cl..L..., y = frac, label = Species), hjust = 1.1, size = 2) +
  theme_bw() +
  scale_x_log10(limits = c(0.0075, 100000)) +
  labs(x = expression(paste('Concentration of Chloride [ mg Cl/L ]')), y = 'Fraction of vertebrates affected')
```



This plot contains both the species sensitivity data for the vertebrates and environmental exposure data where the x-axis is on the log10 scale and the y-axis the species affected. The red dot is the HC5 and the blue dot is the 95th percentile. We can see that the number of species affected by the 95th percentile to be 0%. That is, approximately 95% of the time no species will be affected.