Wilcoxon and Quantiles

Jyotishka Datta

September 8, 2017

Wilcoxon Test with Ties

- Will show you warning message !

```
wilcox.test(c(-2, -2, 2, 2, 5, 5, 5), mu = 3)
```

```
## Warning in wilcox.test.default(c(-2, -2, 2, 2, 5, 5, 5),
## compute exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  c(-2, -2, 2, 2, 5, 5, 5)
## V = 12, p-value = 0.7977
## alternative hypothesis: true location is not equal to 3
```

Suppress Warnings !

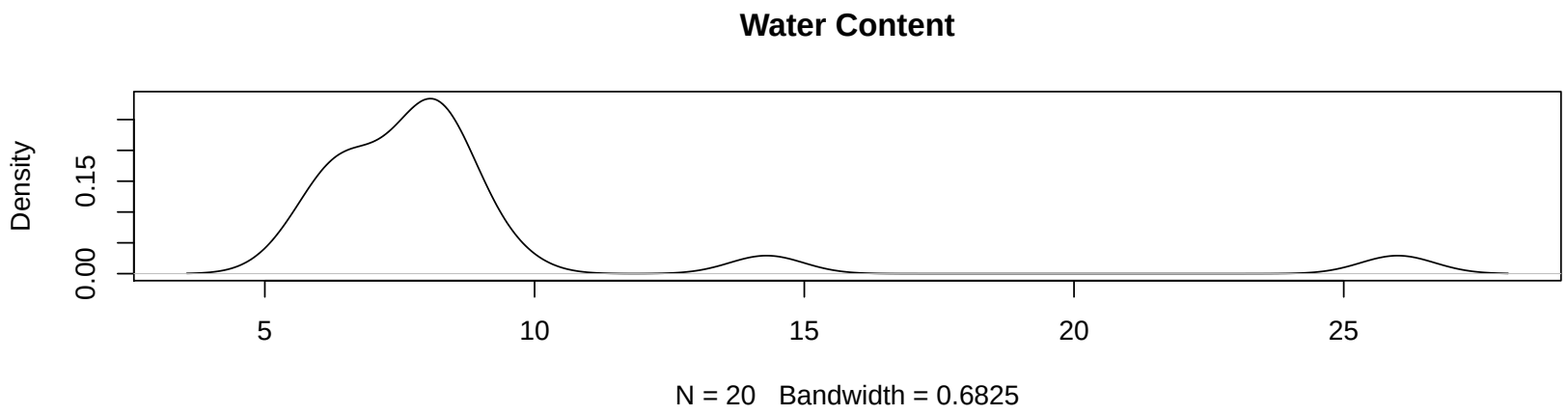
- We can suppress this warning message, but it's not recommended in practice

```
suppressWarnings(wilcox.test(c(-2,-2,2,2,5,5,5),mu = 3))
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  c(-2, -2, 2, 2, 5, 5, 5)
## V = 12, p-value = 0.7977
## alternative hypothesis: true location is not equal to 3
```

Another Example

```
x2 <- c(5.6, 6.1, 6.3, 6.4, 6.5, 6.6, 7.0, 7.5, 7.9, 8.0,8.
## Plot the density
plot(density(x2), main = "Water Content")
```



Wilcoxon Signed Rank Test

```
suppressWarnings(wilcox.test(x2, mu=9, conf.int=TRUE))
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x2
## V = 41, p-value = 0.01774
## alternative hypothesis: true location is not equal to 9
## 95 percent confidence interval:
##  7.150075 8.500071
## sample estimates:
## (pseudo)median
##       7.810093
```

Sign Test

```r
binom.test(sum(x2>9),length(x2),alternative = "two.sided")
```

```
##
##  Exact binomial test
##
## data:  sum(x2 > 9) and length(x2)
## number of successes = 3, number of trials = 20, p-value
## alternative hypothesis: true probability of success is n
## 95 percent confidence interval:
##  0.03207094 0.37892683
## sample estimates:
## probability of success
##                   0.15
```

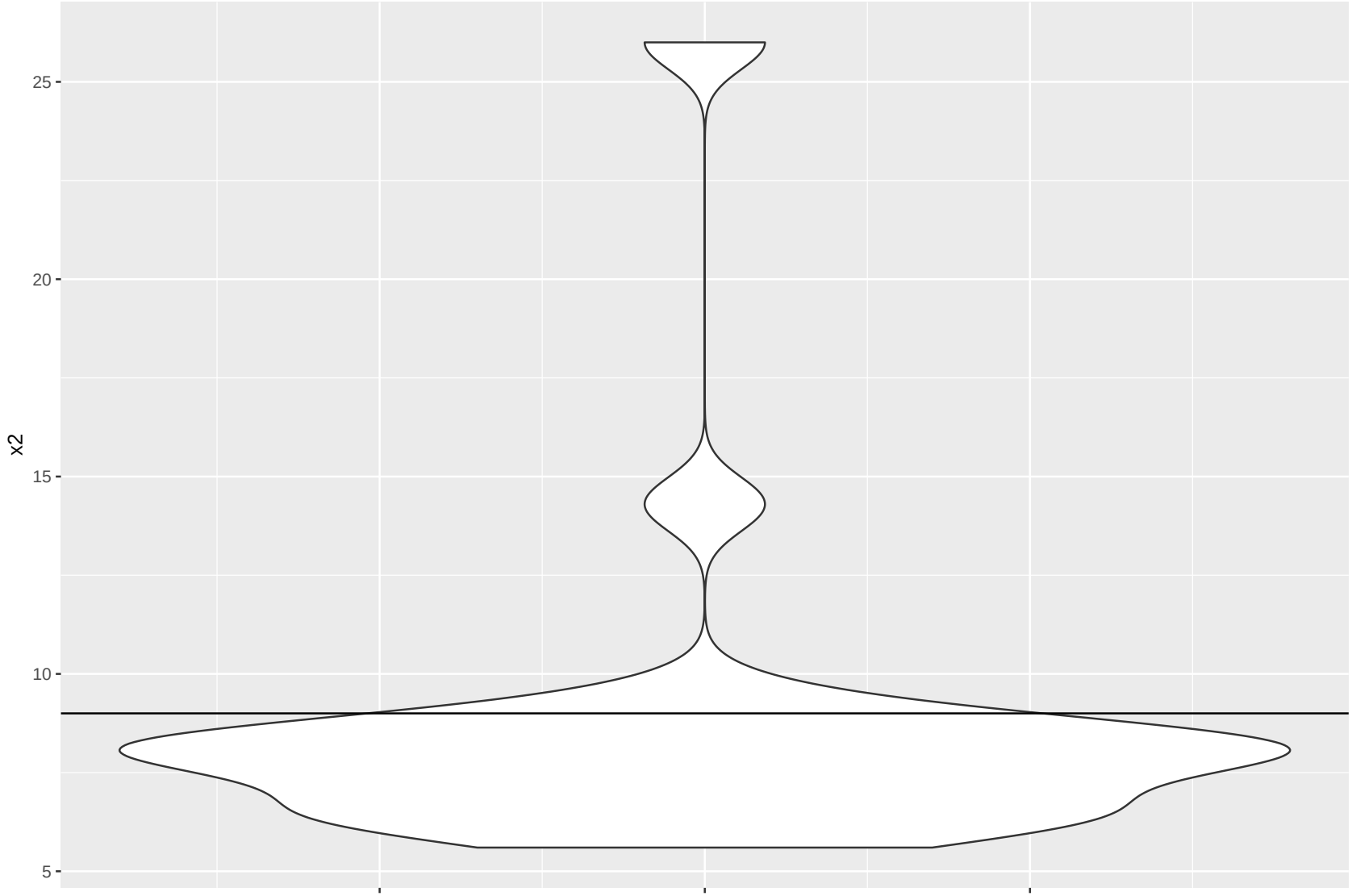
Parametric t-test

```
t.test(x2,alternative = "two.sided",mu=9)
```

```
##
##  One Sample t-test
##
## data:  x2
## t = -0.22154, df = 19, p-value = 0.827
## alternative hypothesis: true mean is not equal to 9
## 95 percent confidence interval:
##   6.701535 10.858465
## sample estimates:
## mean of x
##      8.78
```

Make a nice looking plot

```
library(ggplot2)
qplot(y=x2, x= 1, geom = "violin")+geom_abline(slope=0,inte
```



Quantiles

The ecdf() function in R

Description

Compute an empirical cumulative distribution function, with several methods for plotting, printing and computing with such an "ecdf" object.

Usage

Arguments

x, object

numeric vector of the observations for ecdf; for the methods, an object inheriting from class "ecdf".

...

arguments to be passed to subsequent methods, e.g., plot.stepfun for the plot method.

ylab

label for the y-axis.

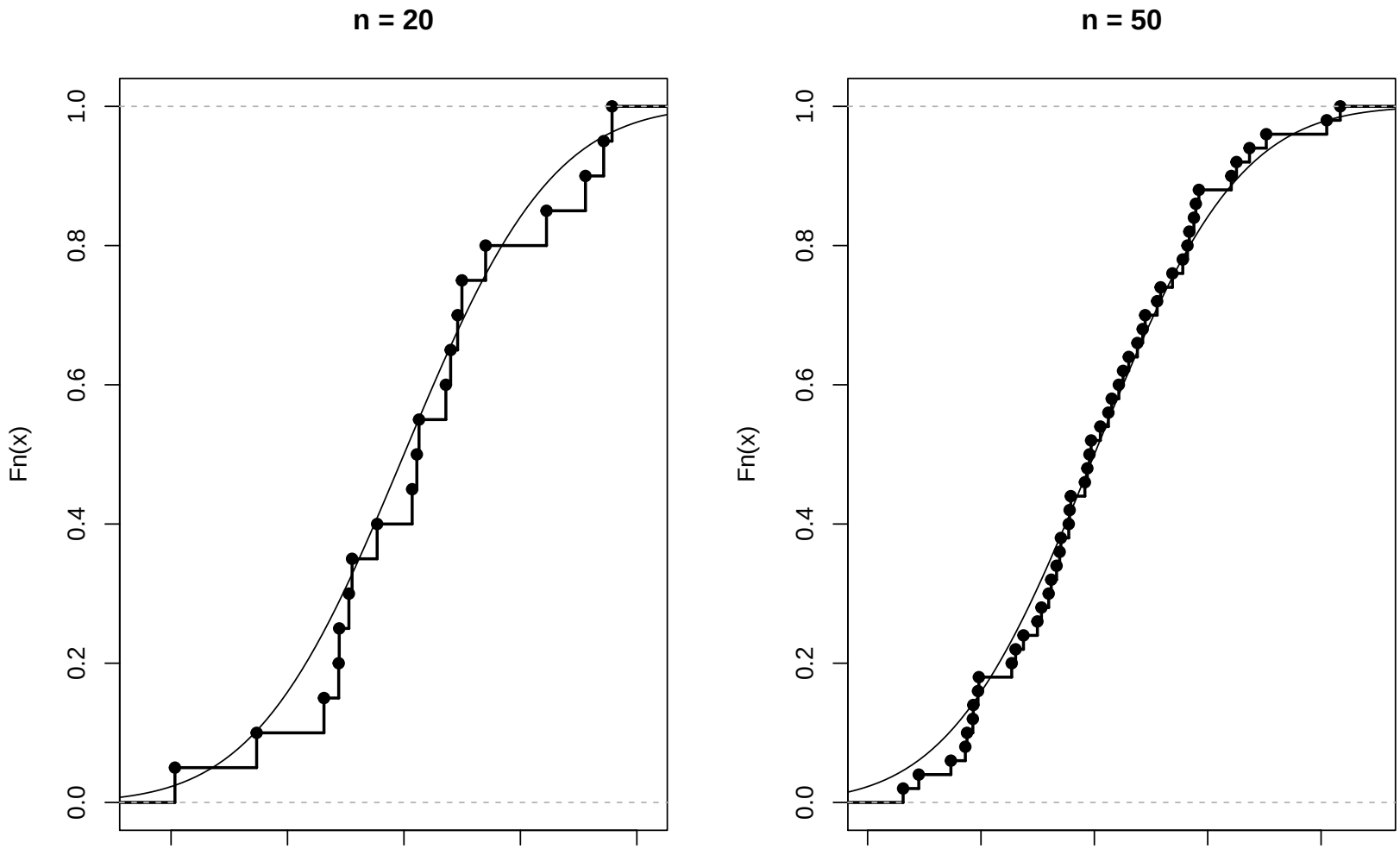
verticals

Convergence

https://jdatta.shinyapps.io/eCDFdemo/

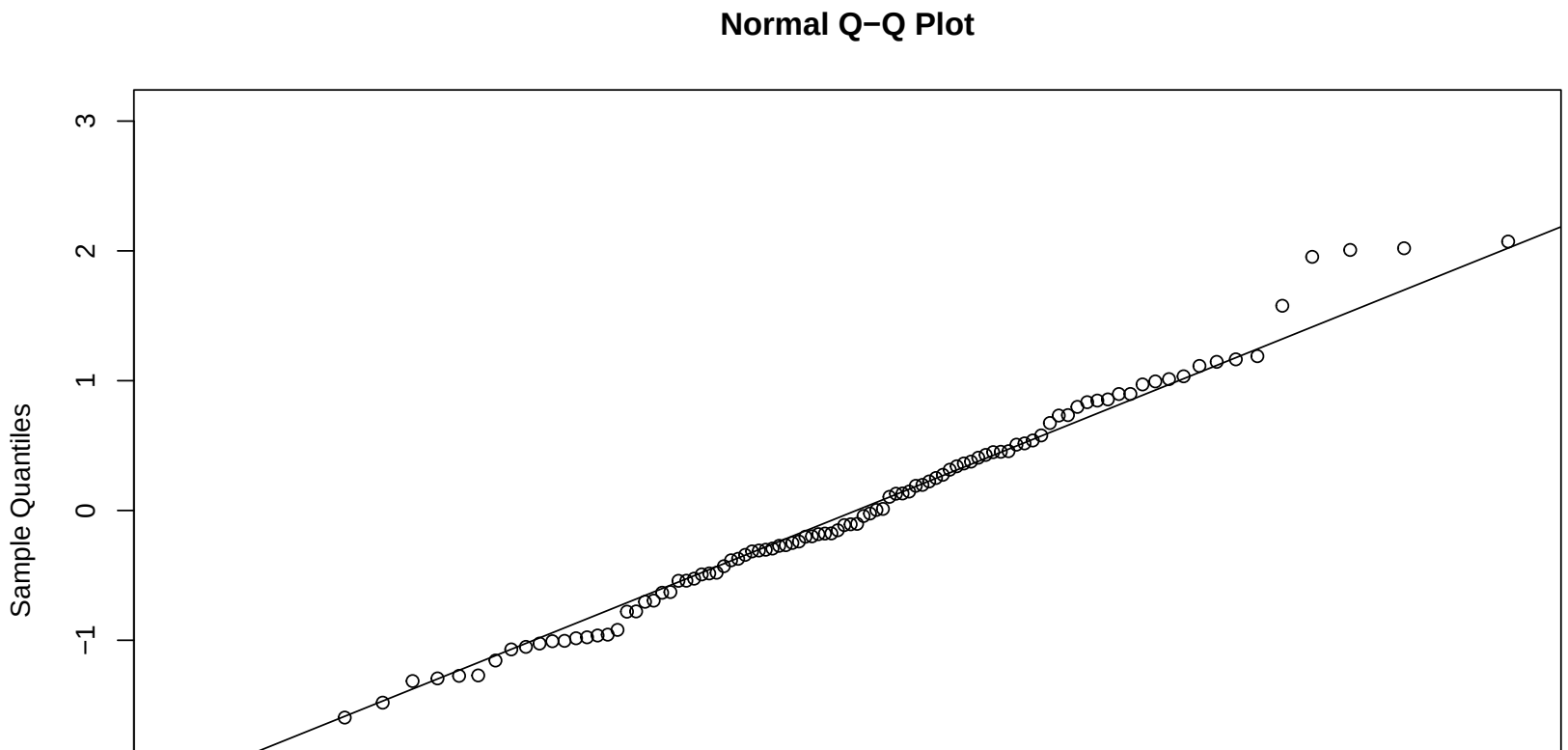
Empirical CDF

```
set.seed(123)
emp1 <- ecdf(rnorm(20)) # E-CDF of 20 samples from N(0,1)
emp2 <- ecdf(rnorm(50)) # E-CDF of 20 samples from N(0,1)
```



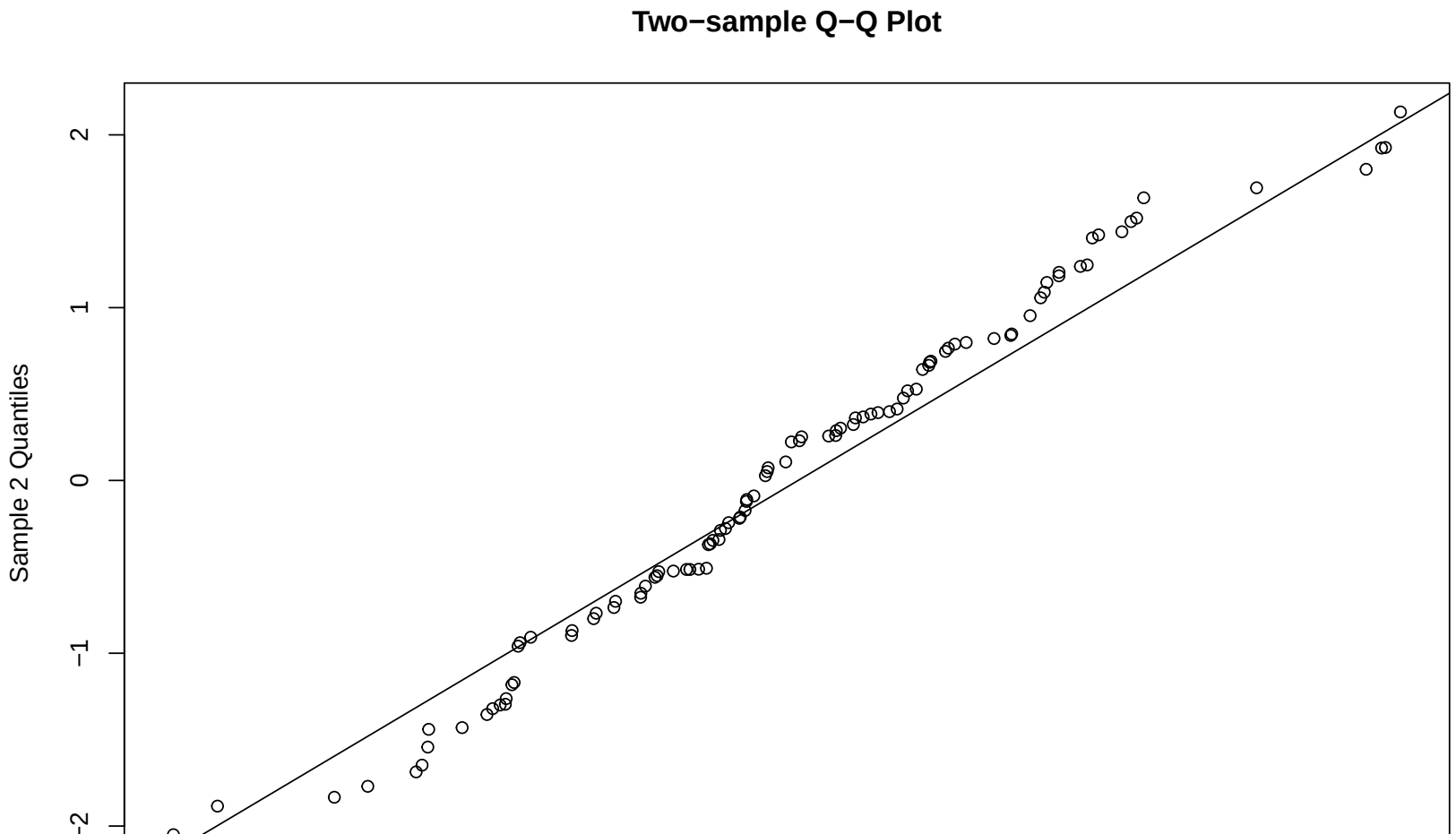
Q-Q Plots (One-sample)

```r
set.seed(12) # Reproducibility
y <- rnorm(100)
qqnorm(y, ylim=c(-3,3), main = "Normal Q-Q Plot",
       xlab = "Theoretical Quantiles", ylab = "Sample Quant
       plot.it = TRUE)
qqline(y, distribution = qnorm)
```



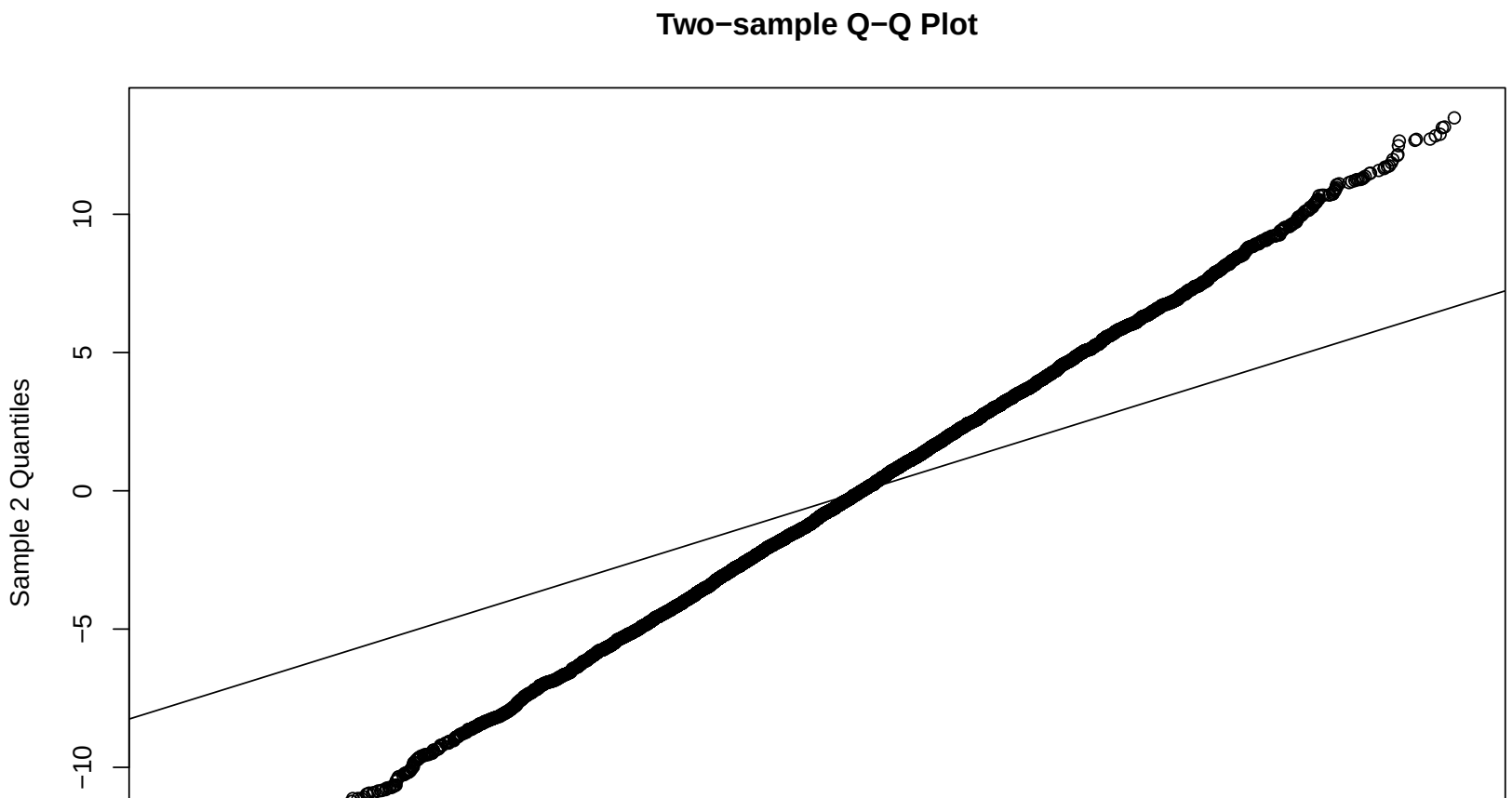
Two samples same

```
z <- rnorm(100)
qqplot(y,z,main = "Two-sample Q-Q Plot",
       xlab = "Sample 1 Quantiles", ylab = "Sample 2 Quanti
abline(0, 1)
```



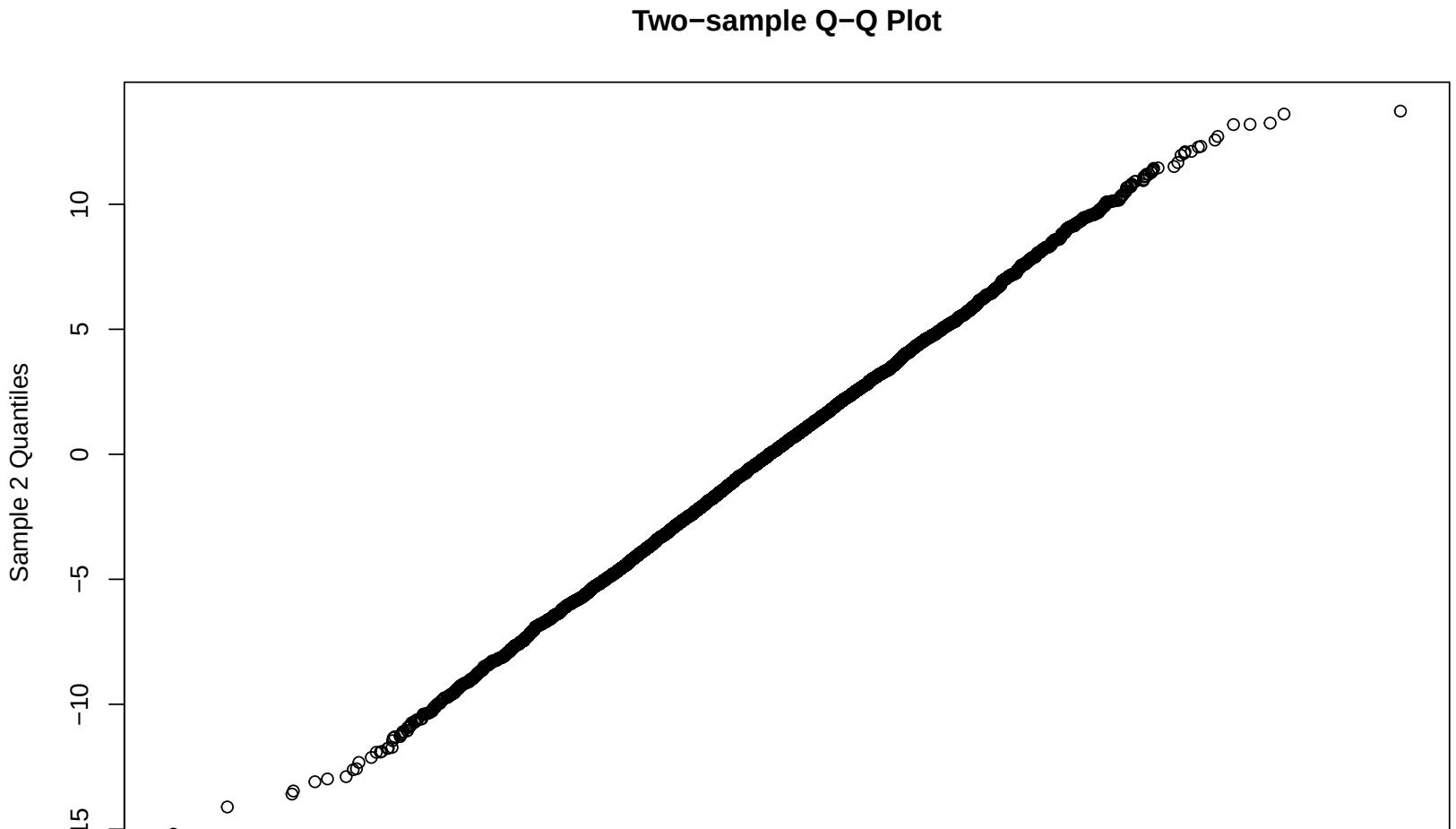
Two samples different

```
x = rnorm(10000,0,2)
y = rnorm(10000,0,4)
qqplot(x,y,main = "Two-sample Q-Q Plot",
       xlab = "Sample 1 Quantiles", ylab = "Sample 2 Quanti
abline(0, 1)
```



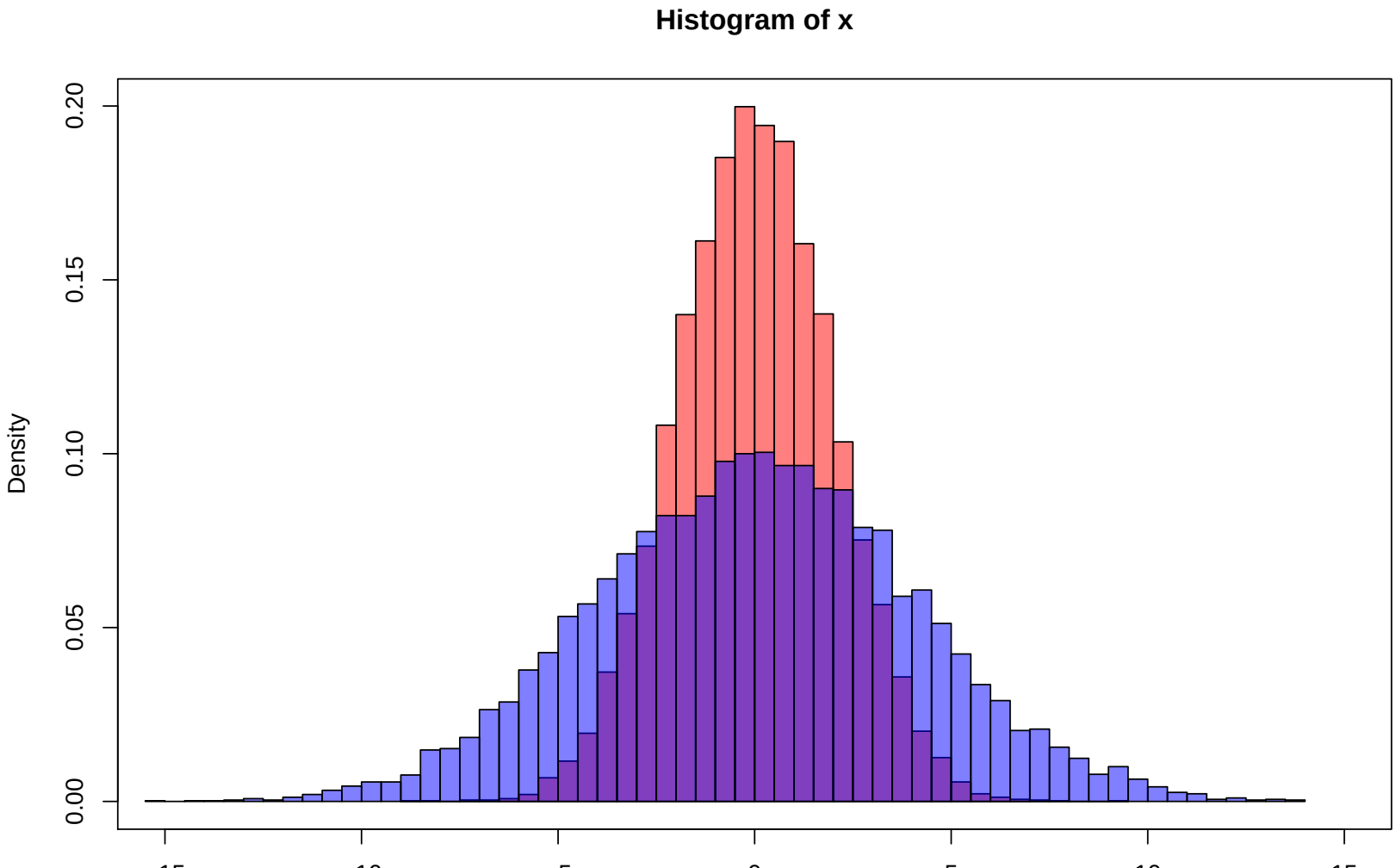
A Common Mistake

```
x = rnorm(10000,0,2)
y = rnorm(10000,0,4)
qqplot(x,y,main = "Two-sample Q-Q Plot",
       xlab = "Sample 1 Quantiles", ylab = "Sample 2 Quanti
```



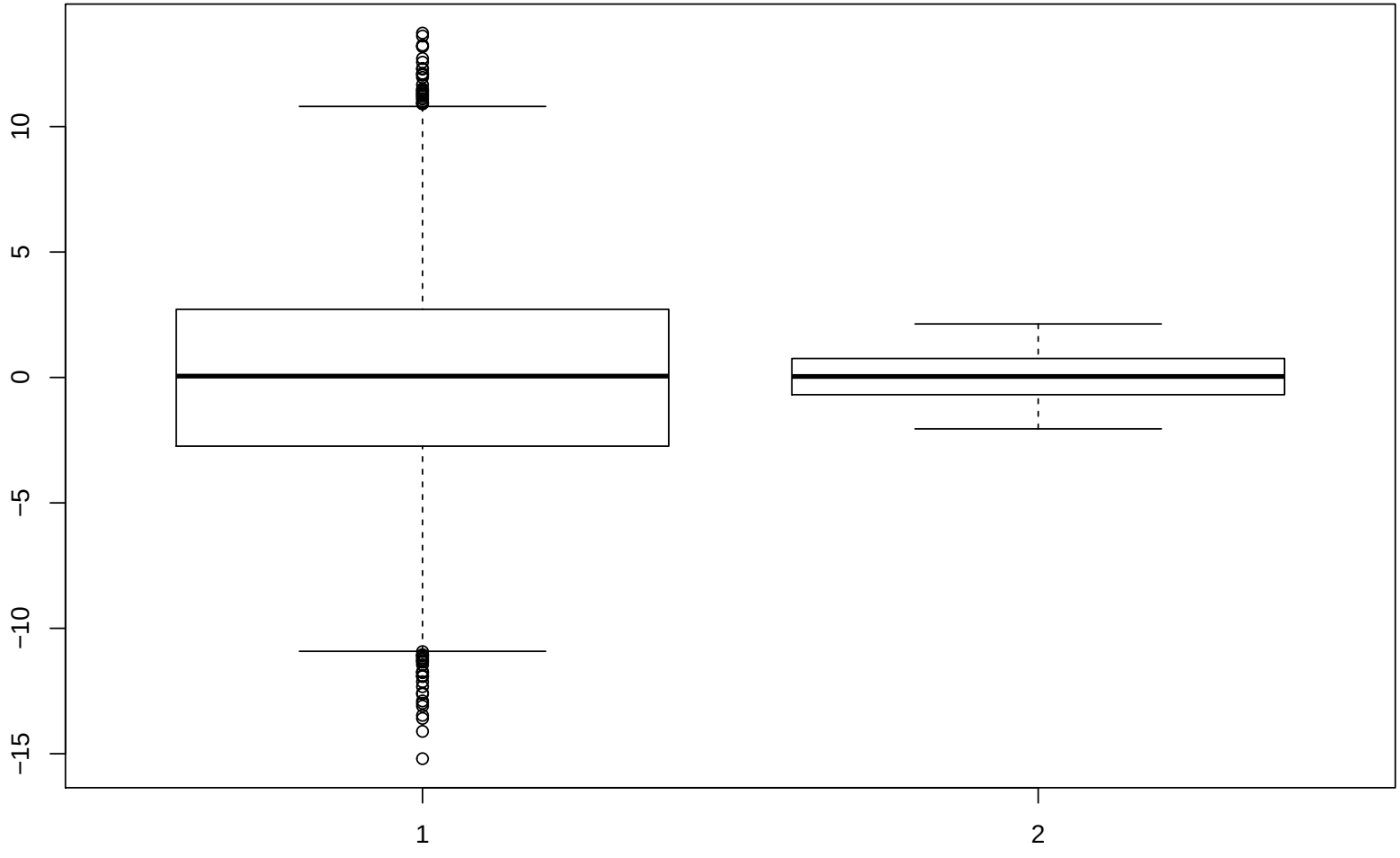
Histograms

```
hist(x,breaks=50,freq=F,col=rgb(1,0,0,0.5),xlim=c(-15,15))
hist(y,breaks=50,freq=F,col=rgb(0,0,1,0.5),add=T)
box()
```



Or, Boxplots

```
boxplot(y,z)
```



Even better, test a hypothesis

```
ks.test(rnorm(20),pnorm)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  rnorm(20)
## D = 0.24294, p-value = 0.1593
## alternative hypothesis: two-sided
```

In our case

```
ks.test(x,y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.1691, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

Next Time

- We will dig deeper into these tests !
- Kolmogorov-Smirnov, Lilliefors, Chi-square GoF etc.
- These are important as they tell you whether you should use NPStats or Normal Stats.