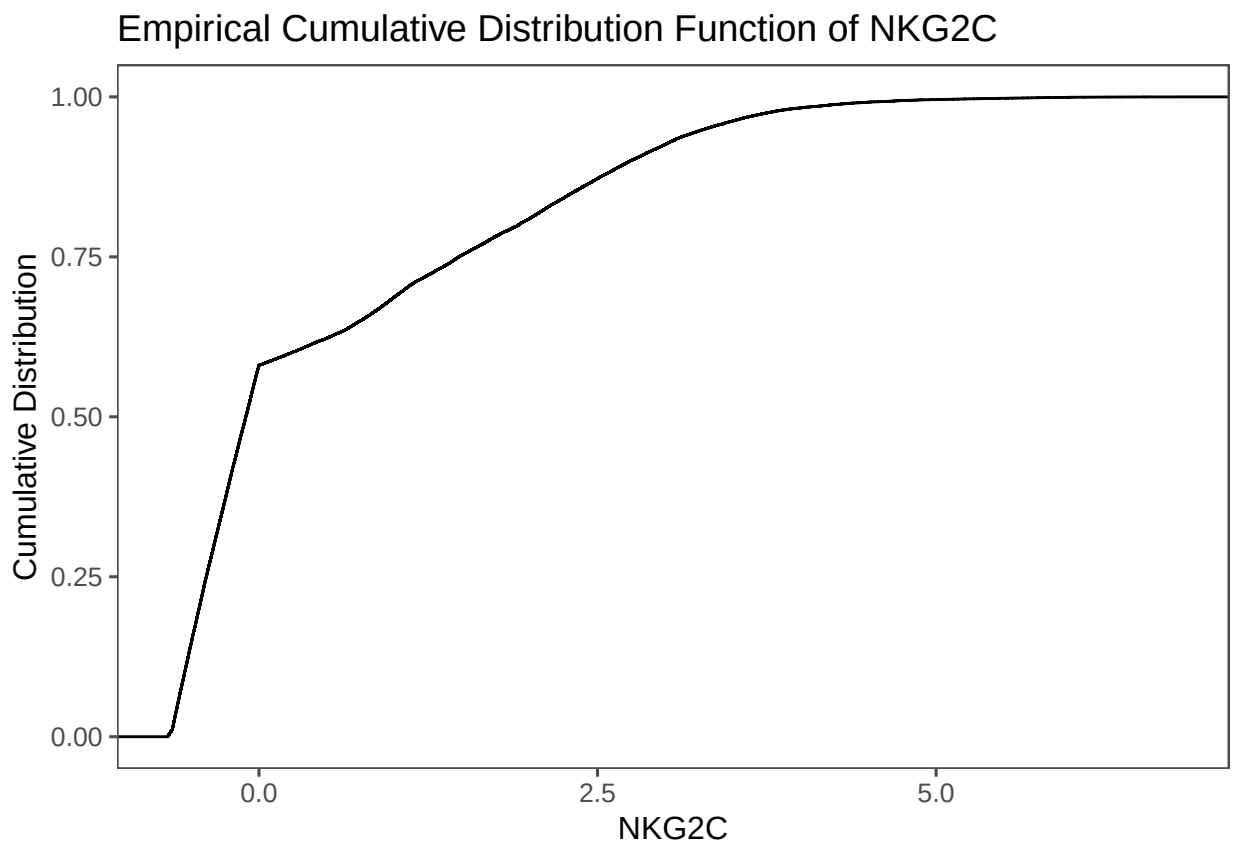
```
library(tidyverse)
```

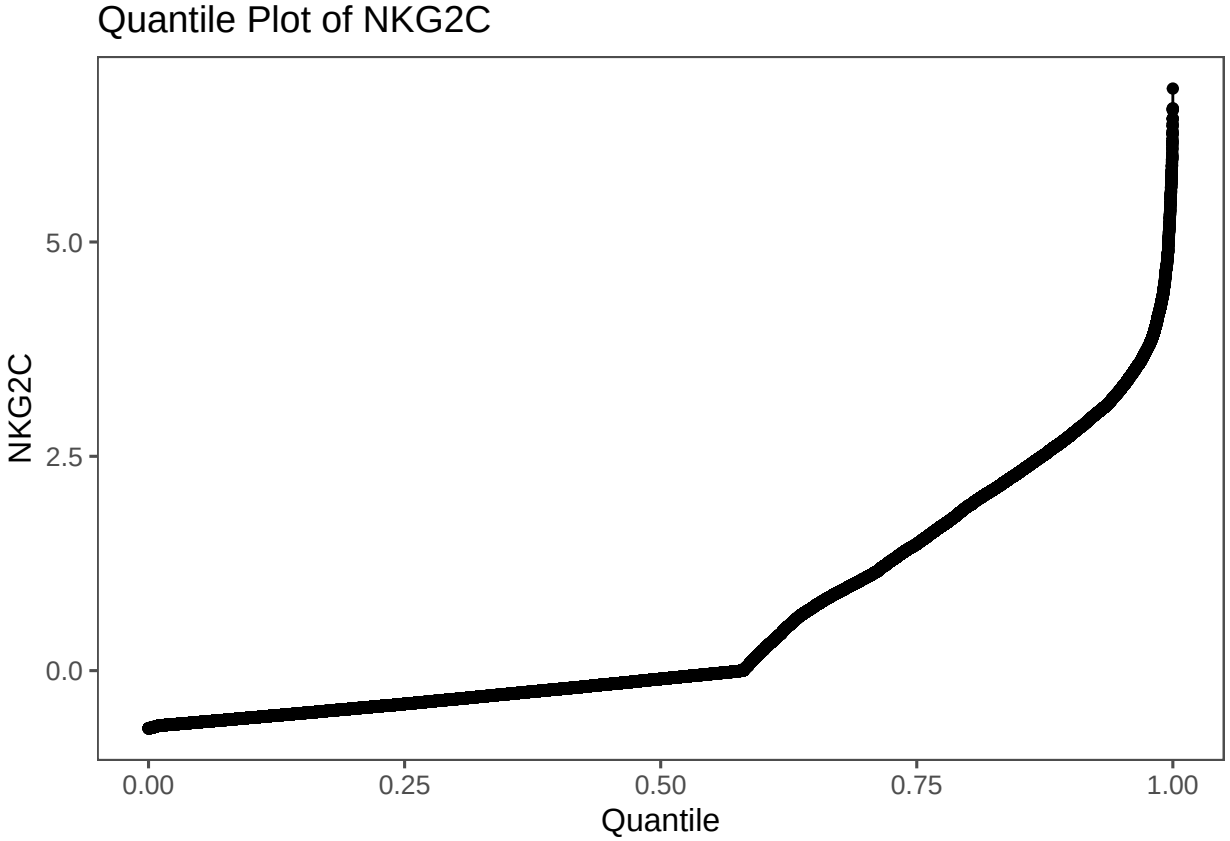
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggthemes)

#1.)
data <- read.csv("cytof_one_experiment.csv")
```

```
#2.)

  #ECDF
  ggplot(data, aes(x=NKG2C)) +
    stat_ecdf() +
    ggtitle("Empirical Cumulative Distribution Function of NKG2C") +
    ylab("Cumulative Distribution") +
    xlab("NKG2C") +
    theme_few()
```



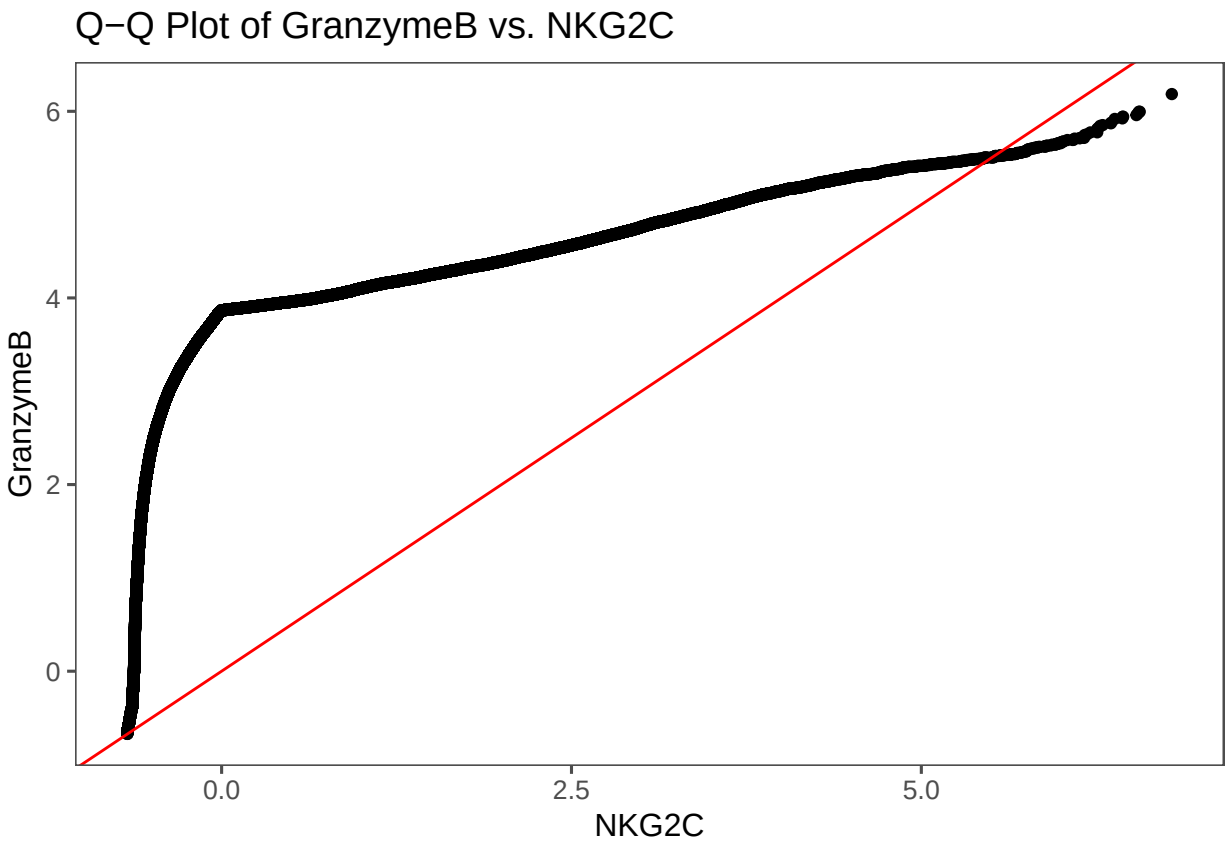
```r
#Quantile
data <- data %>%
          arrange(NKG2C)
n <- nrow(data)
data$f.value <- (0.5:(n - .5))/n
ggplot(data, aes(x=f.value, y=NKG2C)) +
  geom_point() +
  geom_line() +
  ggtitle("Quantile Plot of NKG2C") +
  ylab("NKG2C") +
  xlab("Quantile") +
  theme_few()
```



The two above plots show the univariate distribution of NKG2C. A substantial chunk of NKG2C's density occurs at values 0 and below; the median value (.50 quantile) is less than zero. Additionally, there is a constant accumulation of density from NKG2C=0 to NKG2C=3. This is evidenced by a linear line on the ECDF plot between these two points. Finally, the vast majority of NKG2C values are less than 5 as the ECDF plot has almost reached a value of 1 by this point.

```r
#3.)
qq.df <- as.data.frame(qqplot(data$NKG2C, data$GranzymeB, plot.it = FALSE))
ggplot(qq.df, aes(x=x, y=y)) +
  geom_point() +
  geom_abline(slope=1, intercept=0, col="red") +
  ggtitle("Q-Q Plot of GranzymeB vs. NKG2C")+
  ylab("GranzymeB") +
```

```
  xlab("NKG2C") +
  theme_few()
```



Across the vast majority of quantiles, GranzymeB takes on greater values than NKG2C as the Q-Q plot line is above the identity line. It is only at the lowest and highest quantiles that NKG2C takes on greater values than GranzymeB (where the Q-Q plot line falls below the identity line). This is due to the univariate distribution of NKG2C explored above, where a majority of the density was clumped at values equal to or slightly less than 0.