

DSC_1105_FA1

Frances Aneth Rosales

2024-02-16

Read & Show Data

```
library(readr)

## Warning: package 'readr' was built under R version 4.2.3

filepath <- read_csv("cytof_one_experiment.csv")

## Rows: 50000 Columns: 35
## — Column specification —————
## Delimiter: ","
## dbl (35): NKp30, KIR3DL1, NKp44, KIR2DL1, GranzymeB, CXCR6, CD161, KIR2DS4, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

filepath

## # A tibble: 50,000 × 35
##   NKp30 KIR3DL1 NKp44 KIR2DL1 GranzymeB CXCR6 CD161 KIR2DS4 NKp46 NKG2D
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 0.188 3.62 -0.561 -0.294 2.48 -0.145 -0.315 1.94 4.08 2.62
## 2 1.03 1.70 -0.289 -0.480 3.26 -0.0339 -0.411 3.80 3.73 -0.483
## 3 3.00 6.14 1.90 0.482 4.28 1.95 -0.502 -0.320 4.56 -0.507
## 4 4.30 -0.221 0.243 -0.483 3.35 0.926 3.88 -0.170 4.48 1.93
## 5 -0.439 -0.504 -0.153 0.751 3.19 -0.0589 1.09 -0.0503 0.838 -0.458
## 6 2.09 -0.399 3.46 -0.520 4.35 -0.364 -0.571 -0.450 4.06 3.43
## 7 -0.613 -0.117 -0.451 3.55 1.54 -0.414 0.725 -0.0674 2.67 -0.0838
## 8 -0.341 -0.253 -0.459 2.90 -0.545 -0.612 -0.128 2.79 0.726 2.10
## 9 2.31 -0.364 -0.573 3.84 2.92 2.46 2.25 -0.376 4.33 -0.0168
## 10 3.48 -0.0282 -0.182 4.20 4.74 0.695 5.08 4.19 4.54 2.16
## # ... with 49,990 more rows, and 25 more variables: NKG2C <dbl>, X2B4 <dbl>,
## # CD69 <dbl>, KIR3DL1.S1 <dbl>, CD2 <dbl>, KIR2DL5 <dbl>, DNAM.1 <dbl>,
## # CD4 <dbl>, CD8 <dbl>, CD57 <dbl>, TRAIL <dbl>, KIR3DL2 <dbl>, MIP1b <dbl>,
## # CD107a <dbl>, GM.CSF <dbl>, CD16 <dbl>, TNFa <dbl>, ILT2 <dbl>,
## # Perforin <dbl>, KIR2DL2.L3.S2 <dbl>, KIR2DL3 <dbl>, NKG2A <dbl>,
## # NTB.A <dbl>, CD56 <dbl>, INFg <dbl>
```

```
head(filepath)
```

```
## # A tibble: 6 × 35
##   NKp30 KIR3DL1 NKp44 KIR2DL1 GranzymeB CXCR6 CD161 KIR2DS4 NKp46 NKG2D
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  0.188  3.62 -0.561 -0.294  2.48 -0.145 -0.315  1.94  4.08  2.62
## 2  1.03  1.70 -0.289 -0.480  3.26 -0.0339 -0.411  3.80  3.73 -0.483
## 3  3.00  6.14  1.90  0.482  4.28  1.95 -0.502 -0.320  4.56 -0.507
## 4  4.30 -0.221 0.243 -0.483  3.35  0.926  3.88 -0.170  4.48  1.93
## 5 -0.439 -0.504 -0.153  0.751  3.19 -0.0589  1.09 -0.0503 0.838 -0.458
## 6  2.09 -0.399  3.46 -0.520  4.35 -0.364 -0.571 -0.450  4.06  3.43
## # ... with 25 more variables: NKG2C <dbl>, X2B4 <dbl>, CD69 <dbl>,
## # KIR3DL1.S1 <dbl>, CD2 <dbl>, KIR2DL5 <dbl>, DNAM.1 <dbl>, CD4 <dbl>,
## # CD8 <dbl>, CD57 <dbl>, TRAIL <dbl>, KIR3DL2 <dbl>, MIP1b <dbl>,
## # CD107a <dbl>, GM.CSF <dbl>, CD16 <dbl>, TNFa <dbl>, ILT2 <dbl>,
## # Perforin <dbl>, KIR2DL2.L3.S2 <dbl>, KIR2DL3 <dbl>, NKG2A <dbl>,
## # NTB.A <dbl>, CD56 <dbl>, INFg <dbl>
```

INTERPRET THE DATA

- Upon analyzing the plotting, we have found that the mean of the data is indicated by the mean value of 3.457494 for the “GranzymeB” column. The density map indicates a peak in density within this range of values around 4. The finding derived from the mean calculation is supported by this concentration, which suggests that a sizable fraction of the “GranzymeB” values cluster around the mean. Furthermore, the density plot’s variability indicates that there is spread in the data, showing a range of values, even if the central tendency is around 4. All things considered, this interpretation offers a thorough understanding of the “GranzymeB” value distribution, with a noticeable emphasis on the mean.

```
mean(filepath$GranzymeB)
```

```
## [1] 3.457494
```

Click to Hide/Show Plots

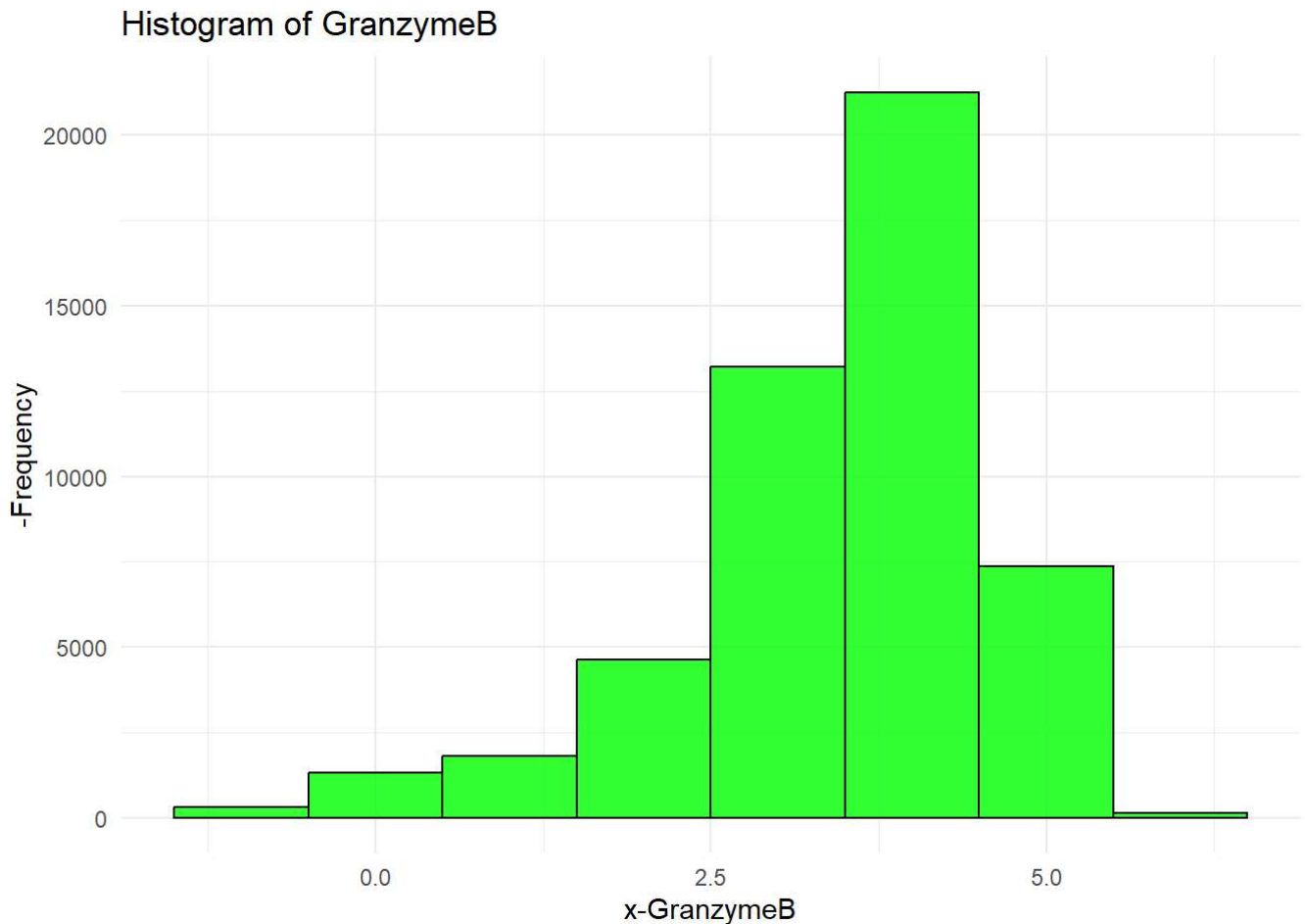
HIDE/SHOW

Histogram

to show the histogram of our desired column “GranzymeB”.

```
library(ggplot2)
```

```
ggplot(filepath, aes(x = GranzymeB)) +  
  geom_histogram(binwidth = 1, fill = "green", color = "black", alpha = 0.8) +  
  labs(title = "Histogram of GranzymeB",  
        x = "x-GranzymeB",  
        y = "-Frequency") +  
  theme_minimal()
```

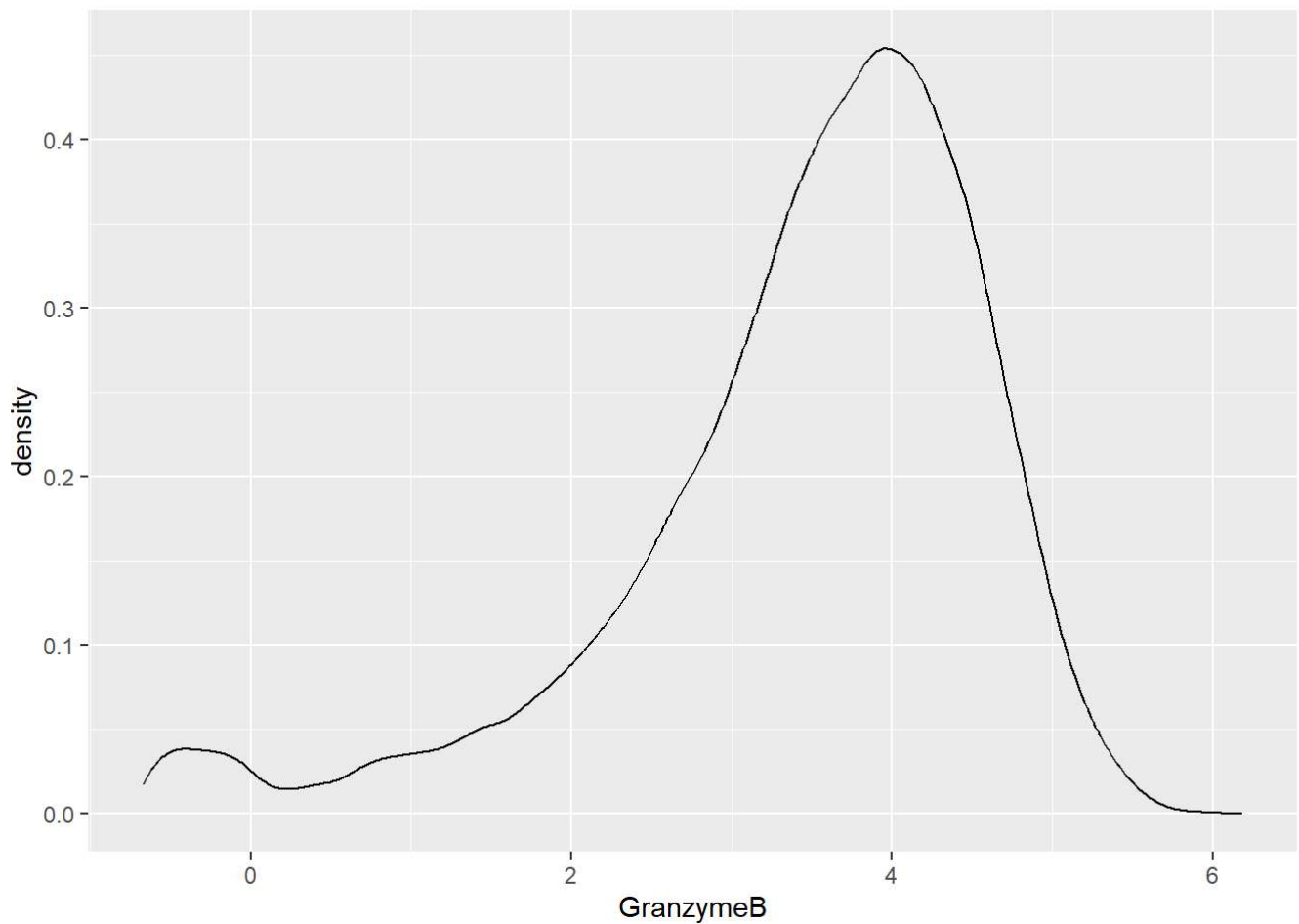


Density Estimates

We can also utilize Density Plotting to analyze the plot of GranzymeB.

```
library(ggplot2)
```

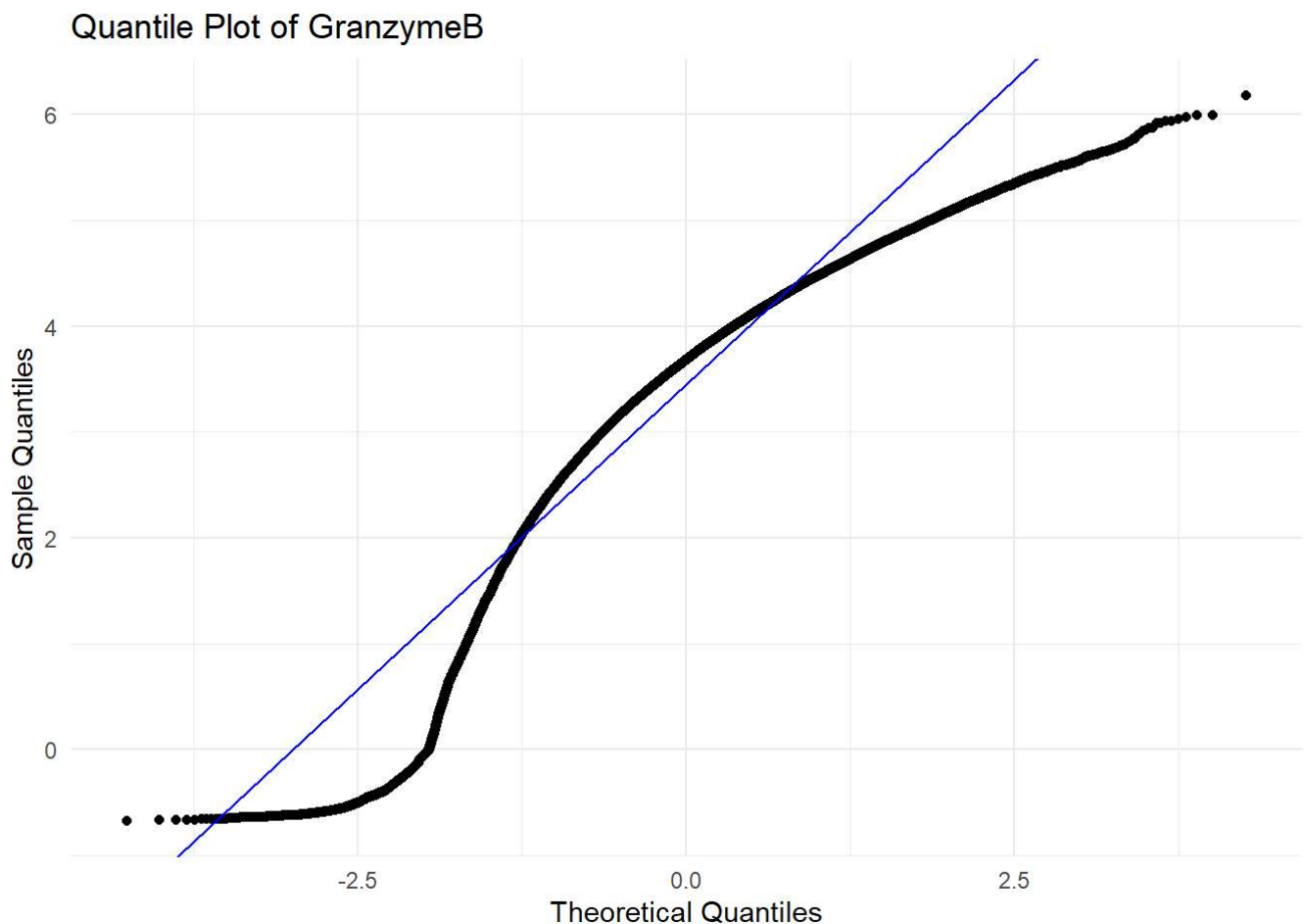
```
ggplot(filepath, aes(x = GranzymeB)) + geom_density()
```



Quantile Plot of GranzymeB

```
library(ggplot2)

ggplot(filepath, aes(sample = GranzymeB)) +
  stat_qq() +
  geom_abline(intercept = mean(filepath$GranzymeB), slope = sd(filepath$GranzymeB), color =
"blue") +
  labs(title = "Quantile Plot of GranzymeB",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()
```



Checking other columns to analyze

```
print(colnames(filepath))
```

```
## [1] "NKp30"      "KIR3DL1"    "NKp44"      "KIR2DL1"
## [5] "GranzymeB"  "CXCR6"      "CD161"      "KIR2DS4"
## [9] "NKp46"      "NKG2D"      "NKG2C"      "X2B4"
## [13] "CD69"       "KIR3DL1.S1" "CD2"        "KIR2DL5"
## [17] "DNAM.1"     "CD4"        "CD8"        "CD57"
## [21] "TRAIL"      "KIR3DL2"    "MIP1b"      "CD107a"
## [25] "GM.CSF"     "CD16"       "TNFa"       "ILT2"
## [29] "Perforin"   "KIR2DL2.L3.S2" "KIR2DL3"    "NKG2A"
## [33] "NTB.A"      "CD56"       "INFg"
```

INTERPRET THE DATA

- The separate distributions for each variable are clearly visible in the NKp30 and NKp44 histograms, as seen in Figure 1.1. The distribution of NKp44 is dispersed between 0 and 1, but the NKp30 histogram shows a range of values around 4. The overlapping zones do not line nicely when attempting to

combine these distributions in the Density Plot at Figure 1.2. The two variables' lack of alignment raises the possibility that they correspond to separate subpopulations or that there are underlying causes influencing each variable's distribution. To comprehend the causes causing the observed discrepancies in the density plots of NKp30 and NKp44, more investigation and analysis could be necessary.

Click to Hide/Show Plots

HIDE/SHOW

```
library(ggplot2)

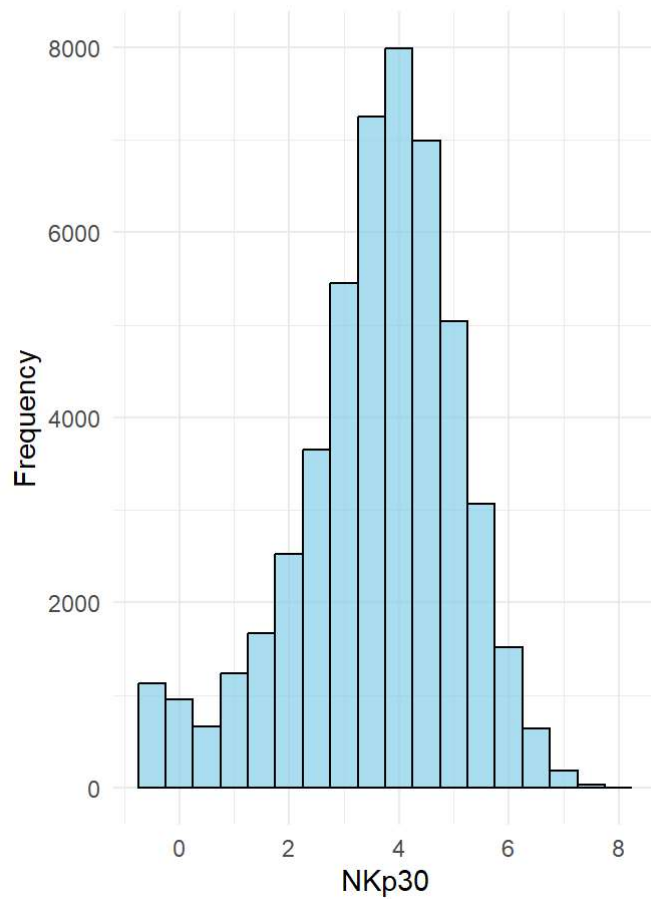
hist_NKp30 <- ggplot(filepath, aes(x = NKp30)) +
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black", alpha = 0.7) +
  labs(title = "Histogram of NKp30",
       x = "NKp30",
       y = "Frequency") +
  theme_minimal()

hist_NKp44 <- ggplot(filepath, aes(x = NKp44)) +
  geom_histogram(binwidth = 0.5, fill = "salmon", color = "black", alpha = 0.7) +
  labs(title = "Histogram of NKp44",
       x = "NKp44",
       y = "Frequency", caption = "Figure 1.1") +
  theme_minimal() +
  theme(plot.caption = element_text(hjust = .2, size = 12, face = "italic"))

combined_plot <- cowplot::plot_grid(hist_NKp30, hist_NKp44, labels = c("A", "B"))

combined_plot
```

A Histogram of NKp30



B Histogram of NKp44

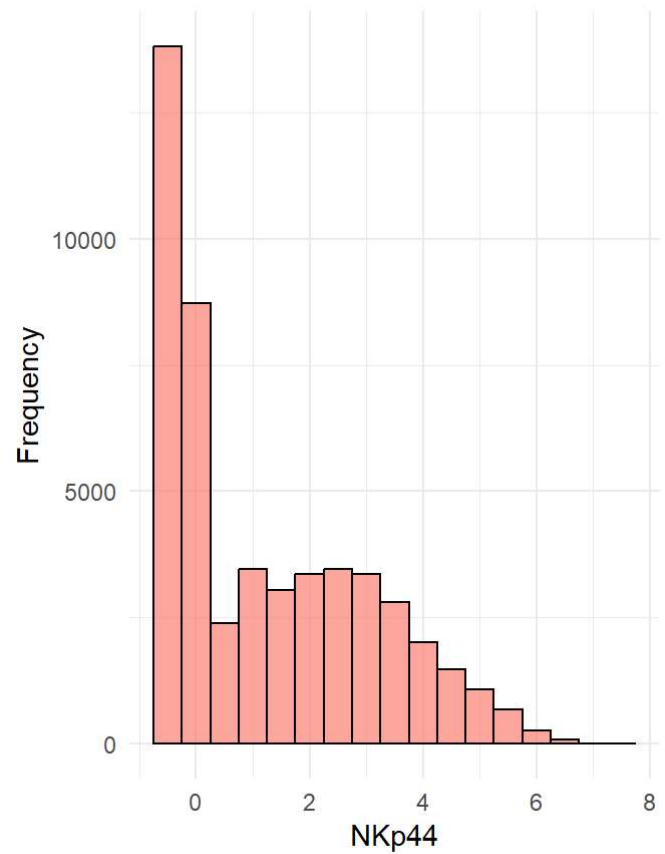


Figure 1.1

```
ggplot(filepath, aes(x = NKp30, fill = "NKp30")) +  
  geom_density(alpha = 0.5) +  
  geom_density(aes(x = NKp44, fill = "NKp44"), alpha = 0.5) +  
  labs(title = "Filled Density Plot of NKp30 and NKp44",  
       x = "Values",  
       y = "Density", caption = "Figure 1.2") +  
  scale_fill_manual(values = c("NKp30" = "blue", "NKp44" = "red")) +  
  theme_minimal() +  
  theme(plot.caption = element_text(hjust = .2, size = 12, face = "italic"))
```

Filled Density Plot of NKp30 and NKp44

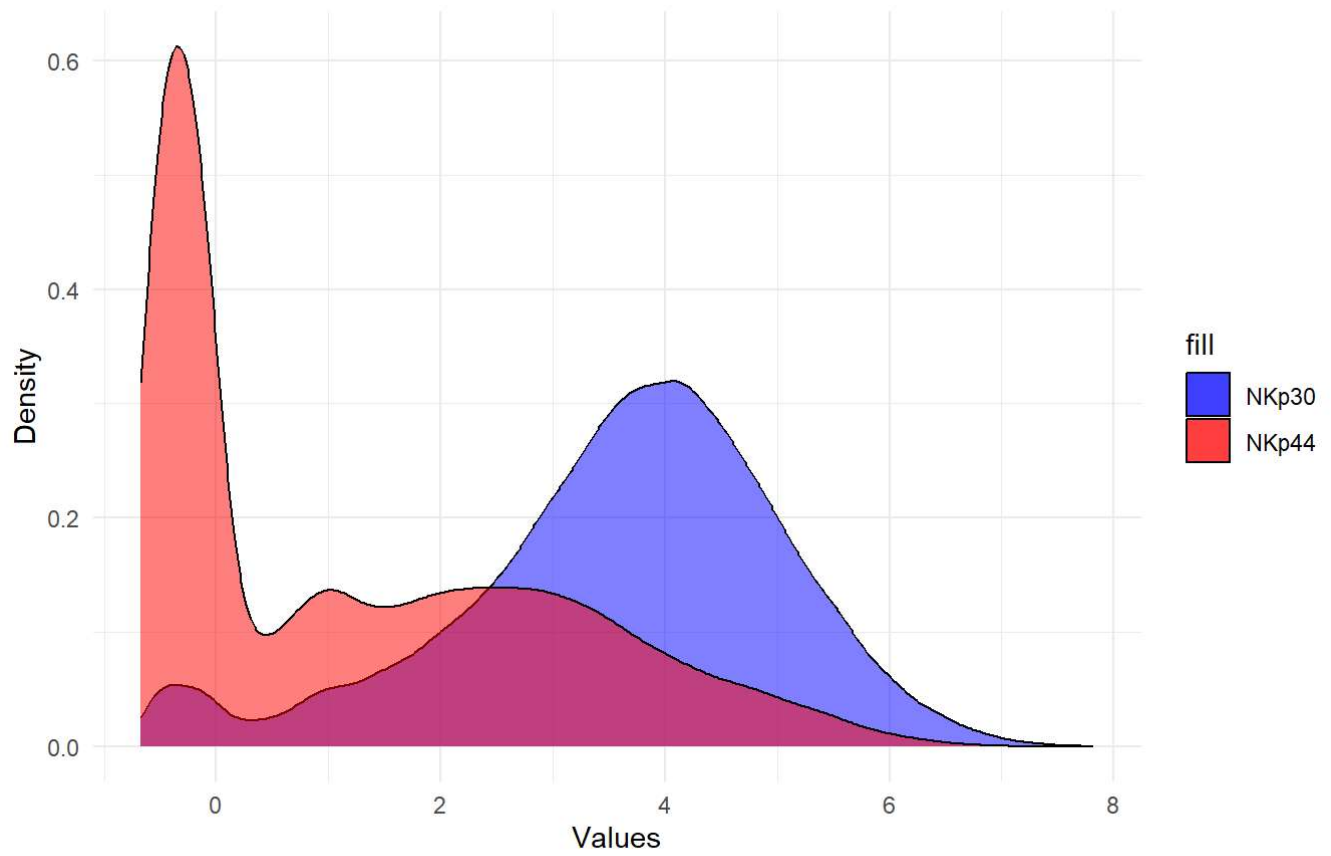


Figure 1.2