

DSC_1105_FA2

Frances Aneth Rosales

2024-02-21

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.
```

```
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>
```

• 1 Using pivot_longer

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
reshaped_data <- filepath %>%  
  pivot_longer(cols = everything(),  
               names_to = "Protein Identity",  
               values_to = "Amount")  
  
print(reshaped_data[, c("Protein Identity", "Amount")])
```

```
## # A tibble: 1,750,000 × 2  
##   `Protein Identity` Amount  
##   <chr>             <dbl>  
## 1 NKp30             0.188  
## 2 KIR3DL1           3.62  
## 3 NKp44            -0.561  
## 4 KIR2DL1          -0.294  
## 5 GranzymeB         2.48  
## 6 CXCR6            -0.145  
## 7 CD161            -0.315  
## 8 KIR2DS4           1.94  
## 9 NKp46             4.08  
## 10 NKG2D            2.62  
## # ... with 1,749,990 more rows
```

Number of ROWS

```
library(dplyr)  
library(tidyr)  
  
reshaped_data <- filepath %>%  
  pivot_longer(cols = everything(),  
               names_to = "Protein Identity",  
               values_to = "Amount")  
  
print(nrow(reshaped_data))
```

```
## [1] 1750000
```

• 2 Using group_by

```
library(dplyr)  
groupby_function <- reshaped_data %>%  
  group_by(`Protein Identity`) %>%  
  summarise(  
    Median = median(Amount),  
    MAD = mad(Amount)  
  )
```

TABLE OF MEDIAN & MAD

```
library(magrittr)
```

```
##  
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:tidyr':  
##  
##   extract
```

```
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.2.3
```

```
##  
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':  
##  
## group_rows
```

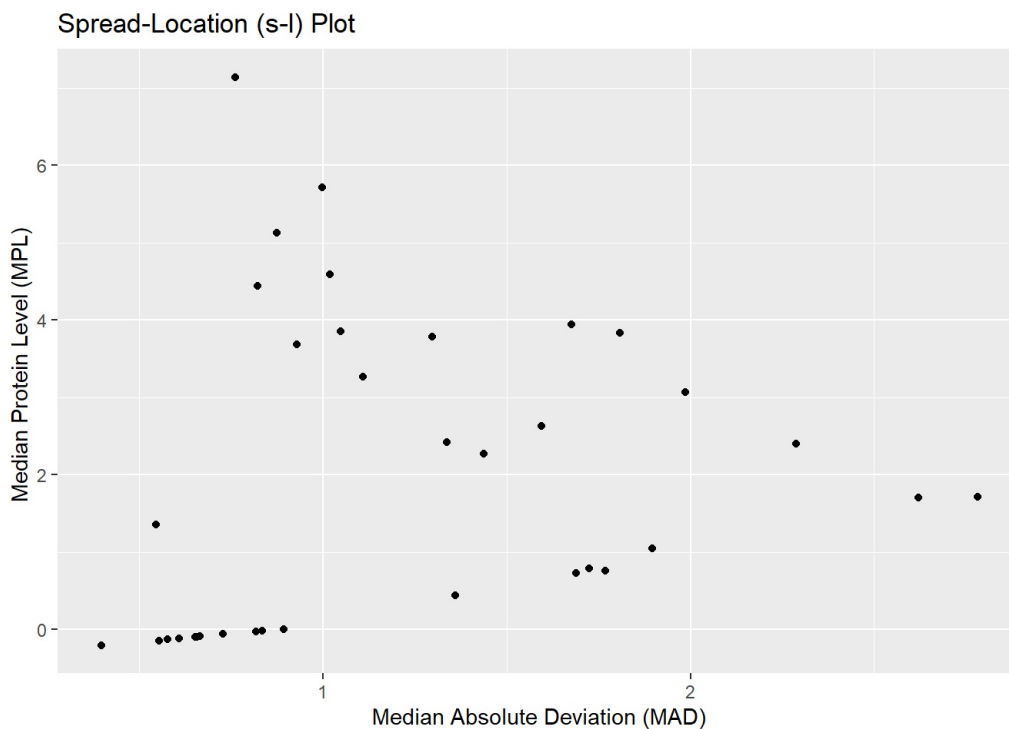
```
kable(groupby_function, "html") %>%  
  kable_styling()
```

Protein Identity	Median	MAD
CD107a	-0.1222997	0.6086976
CD16	5.1229802	0.8744054
CD161	0.7256933	1.6882296
CD2	3.9453789	1.6770427
CD4	-0.2036499	0.3953896
CD56	5.7107903	0.9981358
CD57	3.0709204	1.9868804
CD69	4.5885428	1.0186438
CD8	2.4007181	2.2887240
CXCR6	-0.0581425	0.7266041
DNAM.1	1.3534869	0.5446157
GM.CSF	0.4403750	1.3604700
GranzymeB	3.6828239	0.9278530
ILT2	0.0045176	0.8921193
INFg	2.2654115	1.4365912
KIR2DL1	1.7048951	2.6208342
KIR2DL2.L3.S2	-0.1301092	0.5764354
KIR2DL3	-0.1019813	0.6522079
KIR2DL5	2.4157629	1.3362165
KIR2DS4	1.7102810	2.7821422
KIR3DL1	-0.0212163	0.8345779
KIR3DL1.S1	-0.0927570	0.6635470
KIR3DL2	-0.0338055	0.8167775
MIP1b	3.2699056	1.1092044
NKG2A	3.8344585	1.8083804
NKG2C	-0.0971751	0.6557013
NKG2D	2.6265640	1.5955079
NKp30	3.7795649	1.2968866
NKp44	0.7592514	1.7686227

NKp46	3.8535019	1.0473857
NTB.A	4.4428377	0.8207187
Perforin	7.1410725	0.7604119
TNFa	0.7920363	1.7249212
TRAIL	-0.1442163	0.5539676
X2B4	1.0444406	1.8952282

3 Spread-Location (s-l) Plot

```
library(ggplot2)
ggplot(groupby_function, aes(x = MAD, y = Median)) +
  geom_point() +
  labs(title = "Spread-Location (s-l) Plot",
       x = "Median Absolute Deviation (MAD)",
       y = "Median Protein Level (MPL) ")
```



- Interpretation**

The relationship between the median absolute deviation (MAD) and the median protein level is shown in the spread-location (s-l) plot. The data exhibits a stable degree of variability, with the majority of observations being centered around a particular median value. A steady, predictable distribution of protein levels with few outliers is shown by the s-l plot. This plot offers important insights into the characteristics of the protein data by assisting in the identification of trends and possible outliers.

4 The example_gymnastics_2 DATA:

```
library(dcldata)
library(tidyverse)

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

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

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

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

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

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

## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ forcats 1.0.0 ✓ stringr 1.5.0
## ✓ lubridate 1.9.2 ✓ tibble 3.2.1
## ✓ purrr 1.0.1
## — Conflicts — tidyverse_conflicts() —
## ✖ magrittr::extract() masks tidyr::extract()
## ✖ dplyr::filter() masks stats::filter()
## ✖ kableExtra::group_rows() masks dplyr::group_rows()
## ✖ dplyr::lag() masks stats::lag()
## ✖ purrr::set_names() masks magrittr::set_names()
## i Use the `library(magrittr)` to force all conflicts to become errors

data(example_gymnastics_2)

exempl_func = example_gymnastics_2 %>%
  pivot_longer(cols = -country,
               names_to = "event_year",
               values_to = "Score") %>%
  separate(event_year, into = c("Event", "Year"), sep = "_", convert = TRUE)
```

Click to Hide/Show Data

HIDE/SHOW

TABLE OF GYMNASTICS_2

```
library(magrittr)
library(kableExtra)

kable(exempl_func, "html") %>%
  kable_styling()
```

country	Event	Year	Score
United States	vault	2012	48.132
United States	floor	2012	45.366
United States	vault	2016	46.866
United States	floor	2016	45.999
Russia	vault	2012	46.366
Russia	floor	2012	41.599
Russia	vault	2016	45.733
Russia	floor	2016	42.032

China	vault	2012	44.266
China	floor	2012	40.833
China	vault	2016	44.332
China	floor	2016	42.066