```
rm(list=ls())
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)
data <- 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.
```

```
## 1.)
data <- data %>%
          pivot_longer(names(data), names_to="Protein_Identity", values_to="Protein_Level")
data
```

```
## # A tibble: 1,750,000 x 2
##    Protein_Identity Protein_Level
##    <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
```

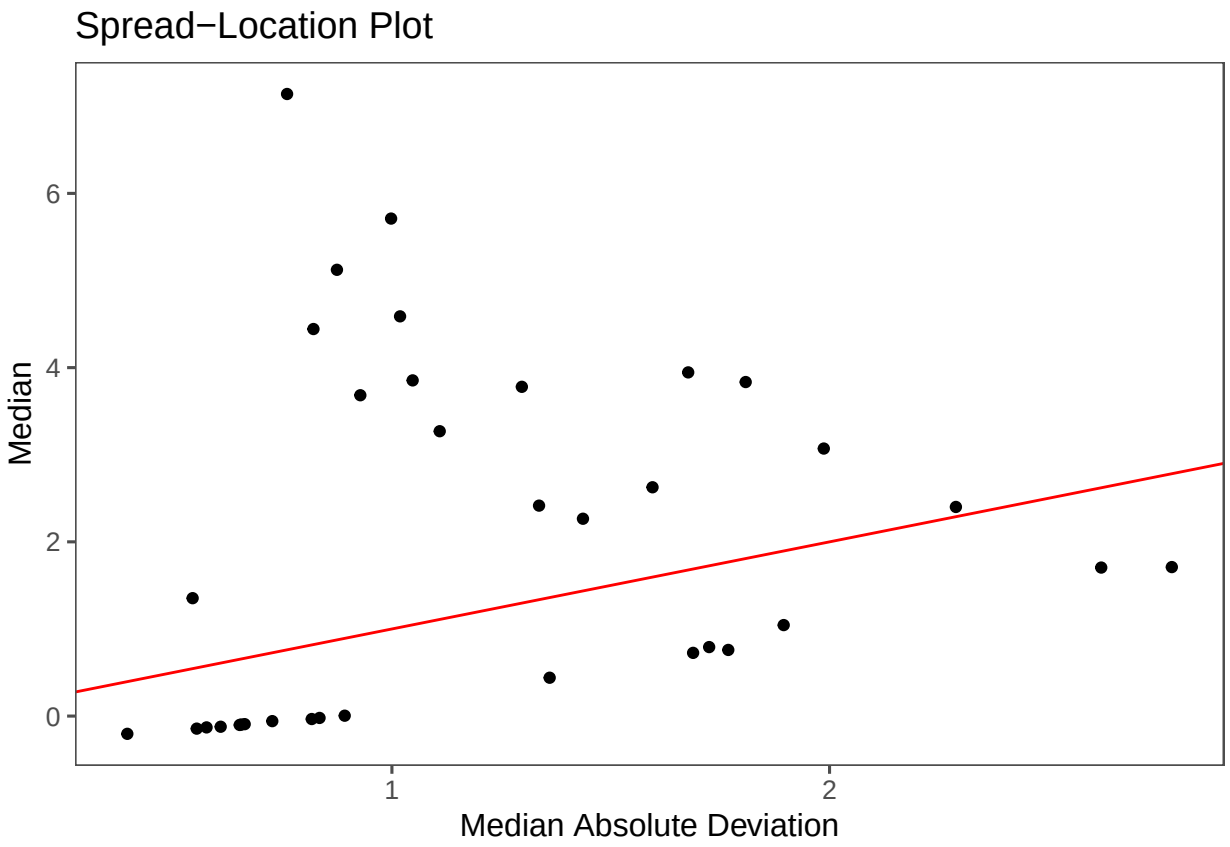
```
nrow(data) == 1750000
```

```
## [1] TRUE
```

```
## 2.)
median_and_mad <- data %>%
                    group_by(Protein_Identity) %>%
                    summarise(Median=median(Protein_Level), MAD=mad(Protein_Level))
median_and_mad
```

```
## # A tibble: 35 x 3
##    Protein_Identity  Median   MAD
##    <chr>              <dbl> <dbl>
##  1 CD107a           -0.122  0.609
##  2 CD16              5.12   0.874
##  3 CD161             0.726  1.69
##  4 CD2               3.95   1.68
##  5 CD4              -0.204  0.395
##  6 CD56              5.71   0.998
##  7 CD57              3.07   1.99
##  8 CD69              4.59   1.02
##  9 CD8               2.40   2.29
## 10 CXCR6            -0.0581 0.727
## # ... with 25 more rows
```

```
## 3.)
ggplot(median_and_mad, aes(x=MAD, y=Median)) +
  geom_point() +
  geom_abline(slope=1, intercept=0, col="red") +
  xlab("Median Absolute Deviation") +
  ylab("Median") +
  ggtitle("Spread-Location Plot") +
  theme_few()
```



```
cor(median_and_mad$MAD, median_and_mad$Median)
```

```
## [1] 0.1542416
```

The spread-location plot is a null plot; there is no clear association between the median absolute deviation and the median. This is noteworthy because the median absolute deviation is just a robust version of the median. The lack of a pattern indicates that the median is unstable in this dataset, and the median absolute deviation should be used as a measure of center.

```
## 4.)
library(dcldata)
data(example_gymnastics_2)
example_gymnastics_2 <- example_gymnastics_2 %>%
  pivot_longer(names(example_gymnastics_2)[-1], names_to="event", values_to="score")
example_gymnastics_2 <- example_gymnastics_2 %>%
  separate(event, into = c("event", "year"))
example_gymnastics_2 <- example_gymnastics_2 %>%
  arrange(desc(score))
example_gymnastics_2
```

```
## # A tibble: 12 x 4
##    country       event year  score
##    <chr>         <chr> <chr> <dbl>
##  1 United States vault 2012   48.1
##  2 United States vault 2016   46.9
##  3 Russia        vault 2012   46.4
##  4 United States floor 2016   46.0
##  5 Russia        vault 2016   45.7
##  6 United States floor 2012   45.4
##  7 China         vault 2016   44.3
##  8 China         vault 2012   44.3
##  9 China         floor 2016   42.1
## 10 Russia        floor 2016   42.0
## 11 Russia        floor 2012   41.6
## 12 China         floor 2012   40.8
```