Gene Expression - Karl 1

Taryn H

2023-03-01

#Inserting Code

data<-read\_csv(here::here("raw-data", "Gene\_Expression.csv"))

## Rows: 88 Columns: 4  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Cell\_Line, Treatment  
## dbl (2): Concentration, Gene\_Expression  
##   
## 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.

data

## # A tibble: 88 × 4  
## Cell\_Line Concentration Treatment Gene\_Expression  
## <chr> <dbl> <chr> <dbl>  
## 1 Wild Type 0 Placebo 5.51  
## 2 Wild Type 1 Placebo 6.41  
## 3 Wild Type 2 Placebo 5.71  
## 4 Wild Type 3 Placebo 7.94  
## 5 Wild Type 4 Placebo 6.87  
## 6 Wild Type 5 Placebo 7.29  
## 7 Wild Type 6 Placebo 10.0   
## 8 Wild Type 7 Placebo 8.85  
## 9 Wild Type 8 Placebo 8.91  
## 10 Wild Type 9 Placebo 9.68  
## # … with 78 more rows

#Clean

skimr::skim\_without\_charts(data)

Data summary

|  |  |
| --- | --- |
| Name | data |
| Number of rows | 88 |
| Number of columns | 4 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 2 |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Cell\_Line | 0 | 1 | 9 | 13 | 0 | 2 | 0 |
| Treatment | 0 | 1 | 7 | 20 | 0 | 2 | 0 |

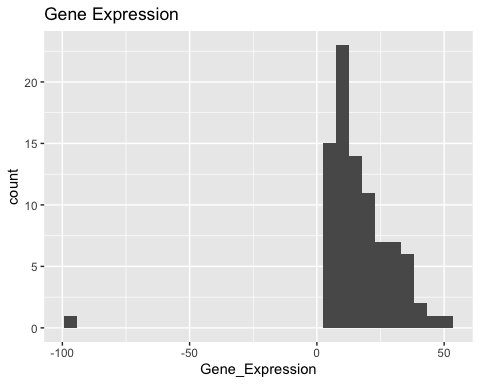
**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Concentration | 0 | 1 | 5.00 | 3.18 | 0 | 2.00 | 5.00 | 8.00 | 10.00 |
| Gene\_Expression | 0 | 1 | 16.14 | 16.45 | -99 | 8.96 | 14.82 | 24.28 | 48.96 |

#Investigate gene expression

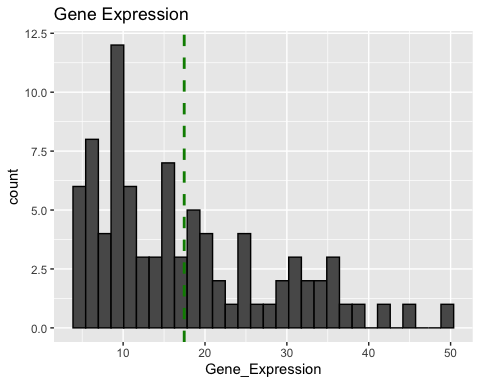
data |> ggplot(aes(x=Gene\_Expression)) + geom\_histogram() + ggtitle("Gene Expression")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#We can see there is a value of -99 which has been entered incorrectly - we will set this to NA  
  
data <- data |> filter(Gene\_Expression>0)  
  
data |> ggplot(aes(x=Gene\_Expression)) + geom\_histogram(col="black") + ggtitle("Gene Expression") + geom\_vline(aes(xintercept=mean(Gene\_Expression)),col="green4",linetype="dashed",size=1) + scale\_color\_brewer(palette="Dark2")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Investigate cell line

class(data$Cell\_Line)

## [1] "character"

table(data$Cell\_Line)

##   
## Cell Type 101 Wild Type   
## 44 43

#Investigate Concentration

class(data$Concentration)

## [1] "numeric"

table(data$Concentration)

##   
## 0 1 2 3 4 5 6 7 8 9 10   
## 8 8 8 8 8 7 8 8 8 8 8

#Investigate Treatment

class(data$Treatment)

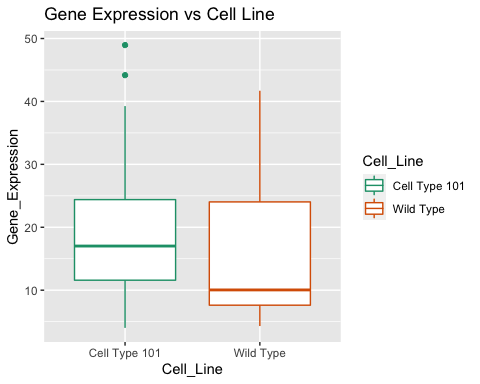
## [1] "character"

table(data$Treatment)

##   
## Activating Factor 42 Placebo   
## 43 44

#Plots

data |> ggplot(aes(x=Cell\_Line, y=Gene\_Expression, col=Cell\_Line)) + geom\_boxplot() + scale\_color\_brewer(palette="Dark2") + ggtitle("Gene Expression vs Cell Line")



data |> ggplot(aes(x=Treatment, y=Gene\_Expression,col=Cell\_Line)) + geom\_boxplot() + scale\_color\_brewer(palette="Dark2") + ggtitle("Gene Expression vs Treatment")



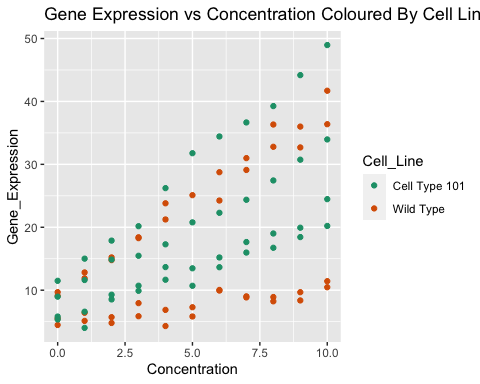
data |> ggplot(aes(x=Concentration, y=Gene\_Expression, col=Treatment)) + geom\_point() + geom\_smooth(method=lm,SE=FALSE)+scale\_color\_brewer(palette="Dark2")+ ggtitle("Gene Expression vs Concentration Coloured By Treatment")

## Warning: Ignoring unknown parameters: SE

## `geom\_smooth()` using formula 'y ~ x'



data |> ggplot(aes(x=Concentration, y=Gene\_Expression, col=Cell\_Line)) + geom\_point() +scale\_color\_brewer(palette="Dark2") + ggtitle("Gene Expression vs Concentration Coloured By Cell Line")



#Make a table

Cell1<-data |>   
 group\_by(Cell\_Line) |>   
 summarise(Mean=mean(Gene\_Expression), Standard\_Deviation=sd(Gene\_Expression))  
  
gt\_data1<-  
 gt(Cell1) |>  
 tab\_header(title="Table 1: Mean of Gene Expression for each Cell")  
gt\_data1

Table 1: Table 1: Mean of Gene Expression for each Cell

| Cell\_Line | Mean | Standard\_Deviation |
| --- | --- | --- |
| Cell Type 101 | 19.20000 | 10.57852 |
| Wild Type | 15.68209 | 10.97189 |

#Make a table 2

Treat1<-data |>   
 group\_by(Treatment) |>  
 summarise(Mean=mean(Gene\_Expression), Standard\_Deviation=sd(Gene\_Expression))  
  
gt\_data2<-  
 gt(Treat1) |>  
 tab\_header(title="Table 2: Mean of Gene Expression for each Treatment")  
gt\_data2

Table 1: Table 2: Mean of Gene Expression for each Treatment

| Treatment | Mean | Standard\_Deviation |
| --- | --- | --- |
| Activating Factor 42 | 24.72419 | 10.421329 |
| Placebo | 10.36341 | 5.036591 |