

Genomics

Marian L. Schmidt, NGS 2015 Tutorial

6/25/2018

```
set.seed(0.11) # Insert your random number here
```

Importing Data

```
metadata <- read.csv('EcoliMetadata.csv') # Load in the data from the data directory!
head(metadata) # This will show us the first 6 rows of the dataframe
```

```
##      sample generation   clade strain      cit      run genome_size
## 1  REL606              0    <NA> REL606 unknown      4.62
## 2 REL1166A            2000 unknown REL606 unknown SRR098028    4.63
## 3  ZDB409             5000 unknown REL606 unknown SRR098281    4.60
## 4  ZDB429            10000      UC REL606 unknown SRR098282    4.59
## 5  ZDB446            15000      UC REL606 unknown SRR098283    4.66
## 6  ZDB458            20000 (C1,C2) REL606 unknown SRR098284    4.63
```

```
str(metadata) # This will show us the structure of the data
```

```
## 'data.frame':   30 obs. of  7 variables:
## $ sample      : Factor w/ 30 levels "CZB152","CZB154",...: 7 6 18 19 20 21 22 23 24 25 ...
## $ generation  : int  0 2000 5000 10000 15000 20000 20000 20000 25000 25000 ...
## $ clade       : Factor w/ 7 levels "(C1,C2)","C1",...: NA 7 7 6 6 1 1 1 2 4 ...
## $ strain      : Factor w/ 1 level "REL606": 1 1 1 1 1 1 1 1 1 1 ...
## $ cit         : Factor w/ 3 levels "minus","plus",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ run         : Factor w/ 30 levels "", "SRR097977",...: 1 5 22 23 24 25 26 27 28 29 ...
## $ genome_size: num  4.62 4.63 4.6 4.59 4.66 4.63 4.62 4.61 4.65 4.59 ...
```

```
mean(metadata$genome_size) # Calculate the mean genome_size
```

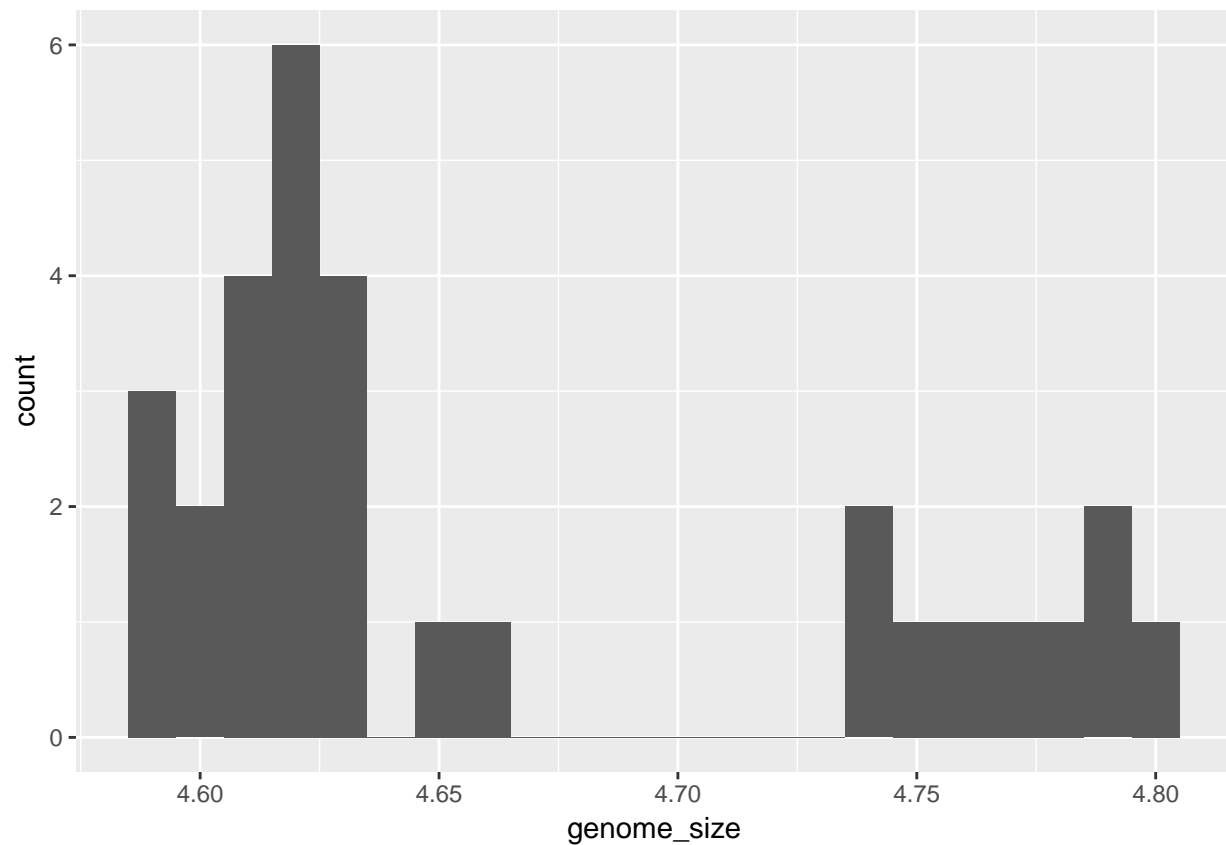
```
## [1] 4.662667
```

Make Some Plots

Plot 1: Let's look at the distribution of the genome size

```
ggplot(metadata, aes(x = genome_size)) +
  geom_bar(stat = "bin", binwidth=0.01) # create a bar plot (histogram) with bins by a genome size of 0

## Warning: `geom_bar()` no longer has a `binwidth` parameter. Please use
## `geom_histogram()` instead.
```



Plot 2: Looking at all of the genome sizes for each strain

```
ggplot(metadata, aes(x = sample, y= genome_size, color = generation, shape = cit)) +
  geom_point(size = rel(3.0)) + # we are going to make points
  theme(axis.text.x = element_text(angle=45, hjust=1)) # x-axis text on a 45 degree angle
```