firstproject

Sean Dube

6/29/2020

First Project for GITHUB using IRIS dataset

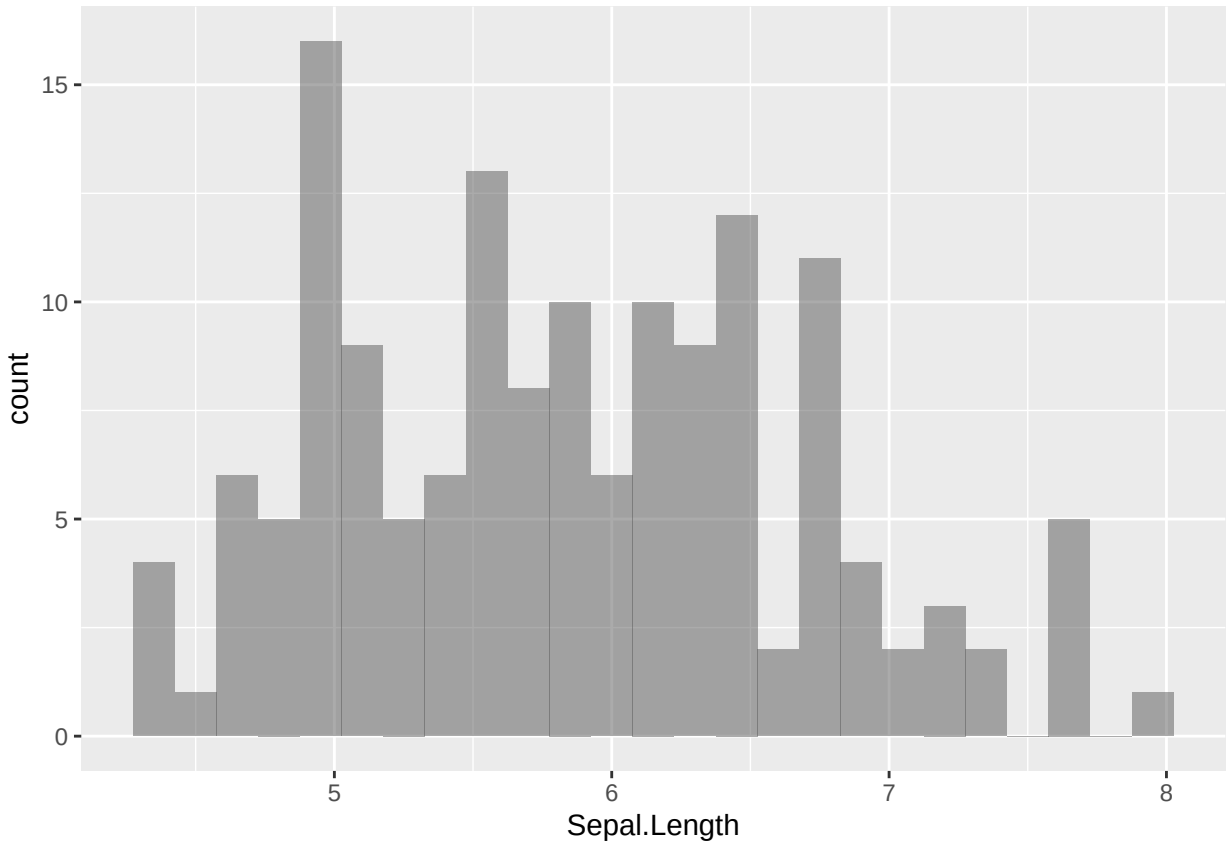
Goal:

1. Use R to analyze the IRIS dataset and answer simple questions.
2. Upload to GITHUB for future use

Analyze IRIS Dataset

Describe the distributions for each variable.

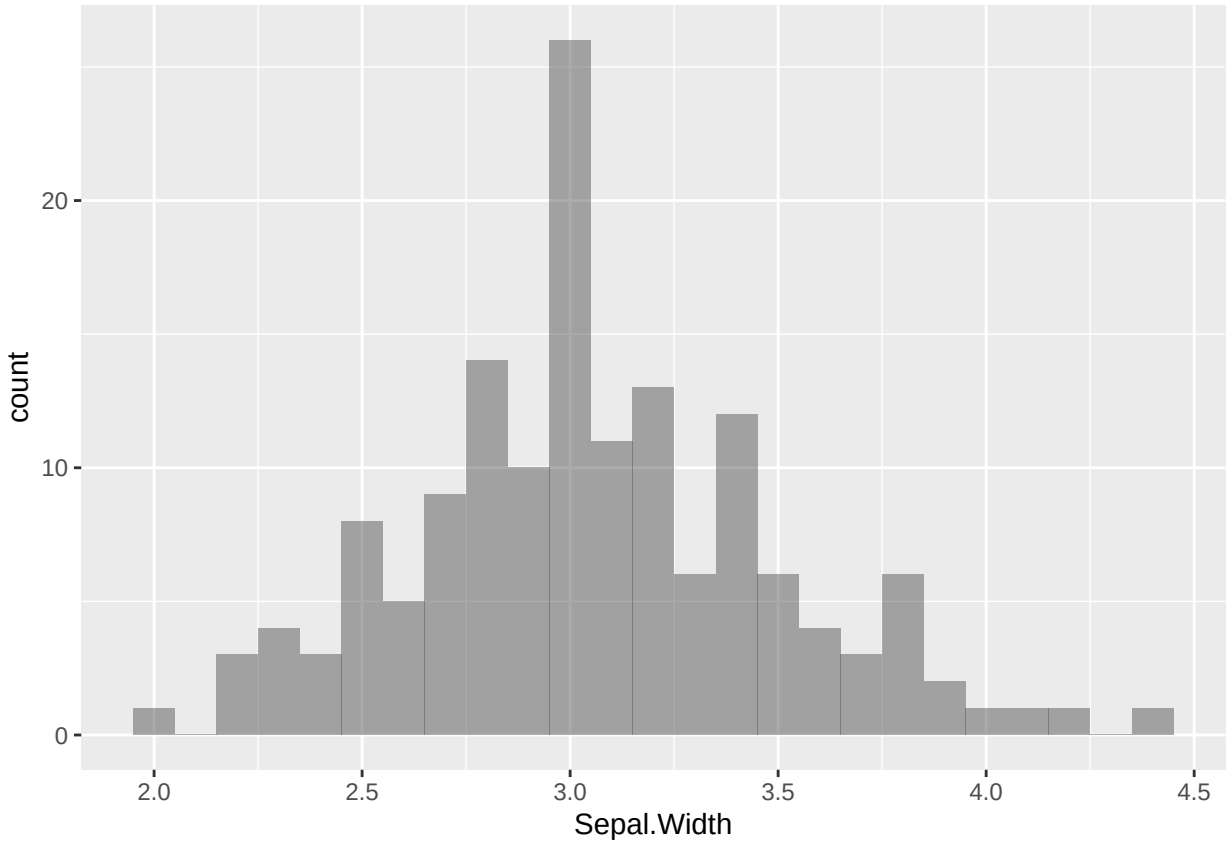
```
gf_histogram(~Sepal.Length, data = iris)
```



```
favstats(~Sepal.Length, data = iris)
```

```
##  min  Q1 median  Q3 max     mean        sd   n missing
##  4.3 5.1    5.8 6.4 7.9 5.843333 0.8280661 150       0
```

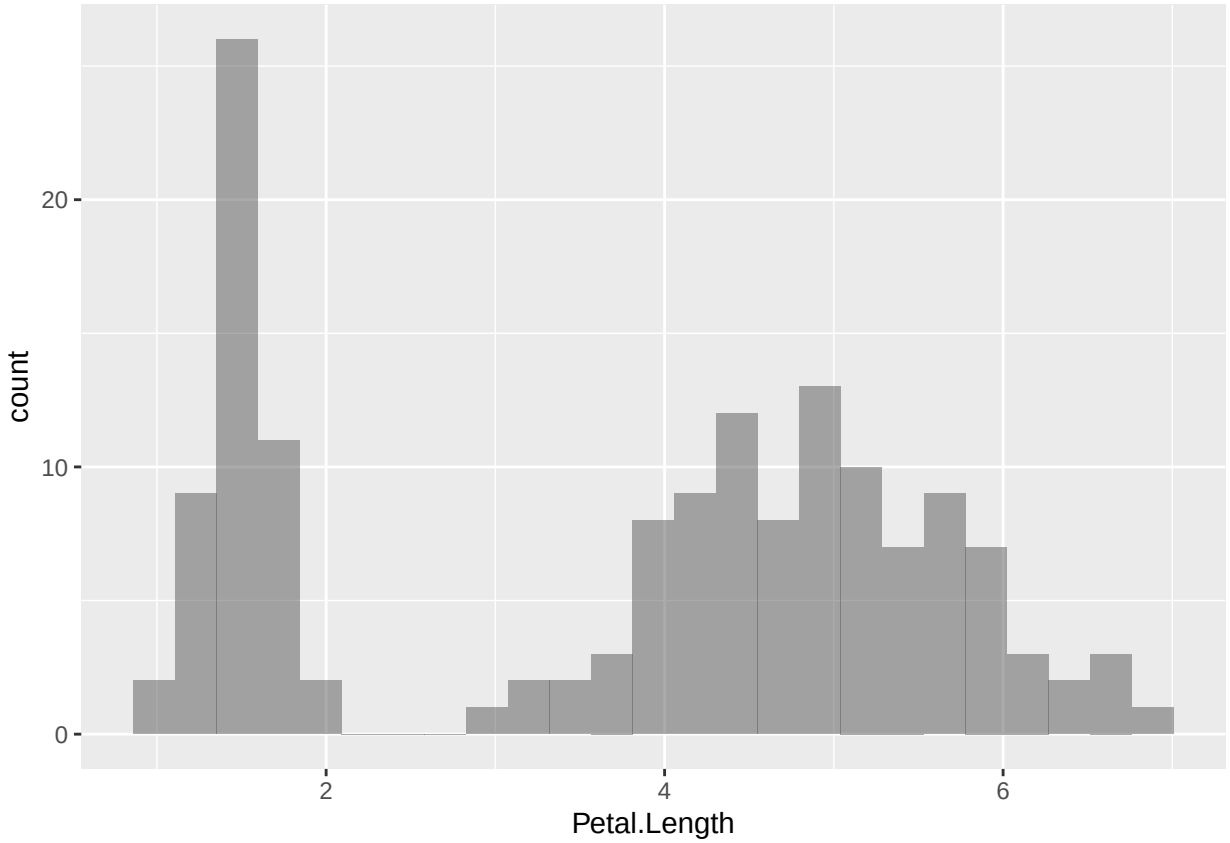
```
gf_histogram(~Sepal.Width, data = iris)
```




```
favstats(~Sepal.Width, data = iris)
```

```
##  min  Q1 median  Q3 max     mean        sd   n missing
##    2 2.8      3 3.3 4.4 3.057333 0.4358663 150       0
```

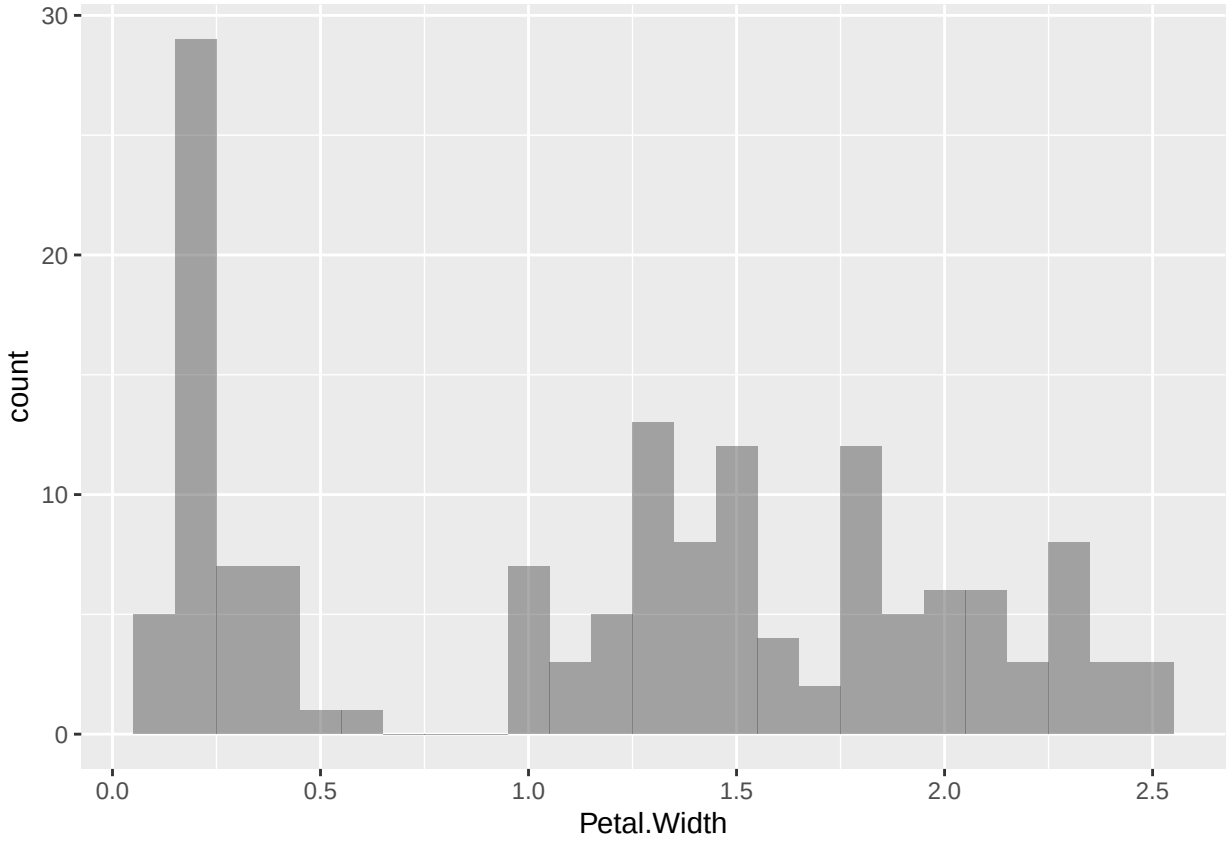
```
gf_histogram(~Petal.Length, data = iris)
```




```
favstats(~Petal.Length, data = iris)
```

```
##  min  Q1 median  Q3 max  mean       sd   n missing
##    1 1.6   4.35 5.1 6.9 3.758 1.765298 150       0
```

```
gf_histogram(~Petal.Width, data = iris)
```




```
favstats(~Petal.Width, data = iris)
```

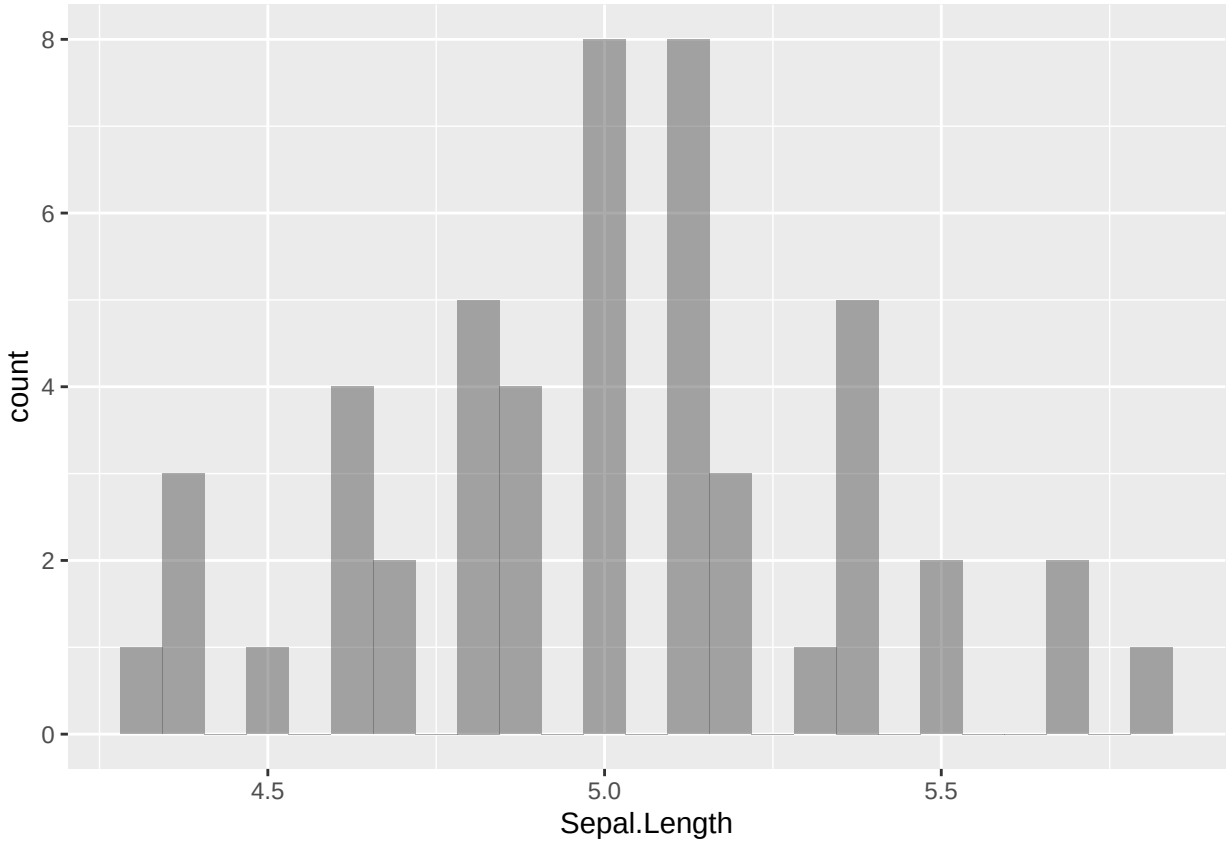
```
##  min  Q1 median  Q3 max     mean        sd   n missing
##  0.1 0.3    1.3 1.8 2.5 1.199333 0.7622377 150       0
```

> Answer: Sepal Length seems to have a right skewed distribution with a median of 5.8cm and an IQR of .5cm. Median and IQR are the correct measure for center and spread since we have skewed distribution. Sepal Width seems to approximately normally distributed with a mean of 3.0573cm and a standard deviation of .4358663cm. Since this distribution is approximately normal we are able to use mean and standard deviations for measures of center and spread. Pedal Length has a right skew with a median of 4.31cmand an IQR of 3.5cm. Lastly, Pedal Width has a right skew with a median of 1.3cm and an IQR of 1.5cm.

Group the flowers by species and then describe the distributions of each variable.

Distributions of each variable for Setosa Species

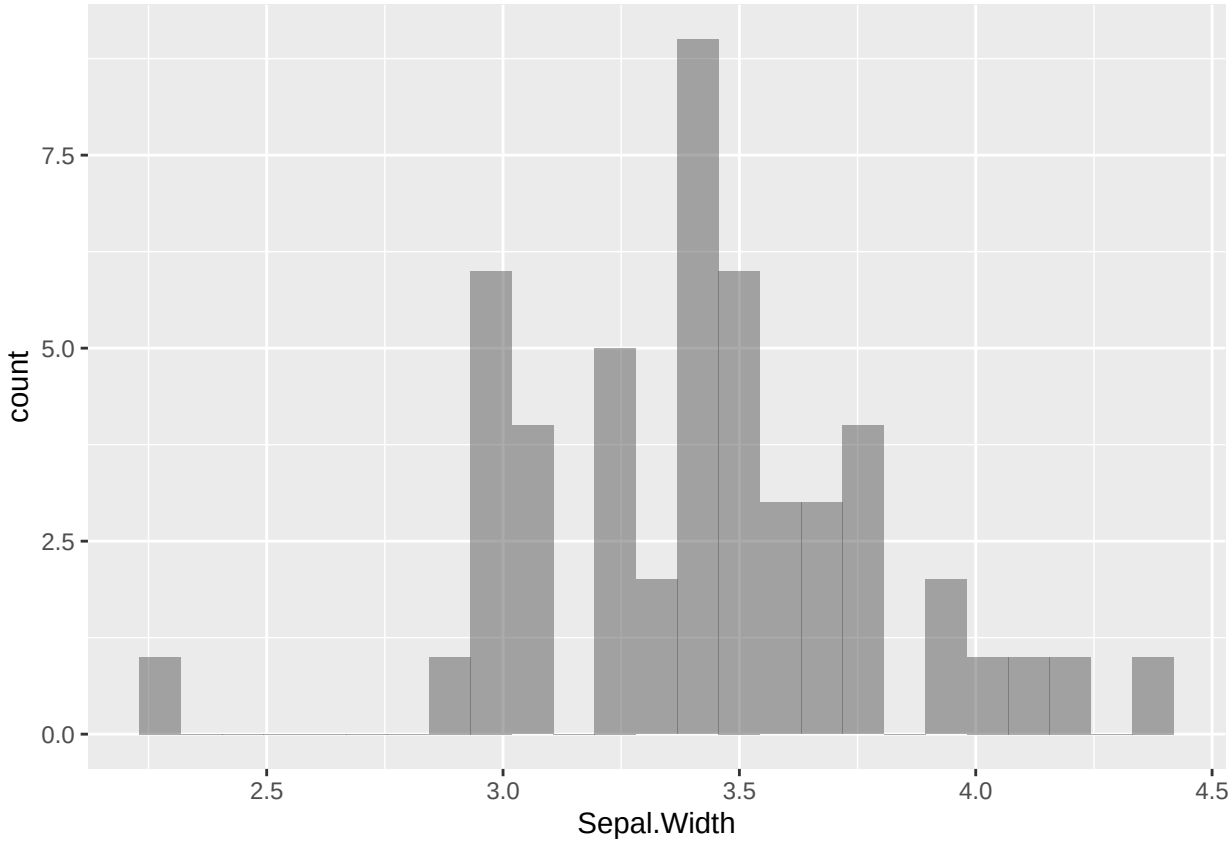
```
setosa <- filter(iris, Species == "setosa")
gf_histogram(~Sepal.Length, data = setosa)
```




```
favstats(~Sepal.Length, data = setosa)
```

```
##  min  Q1 median  Q3 max  mean        sd  n missing
##  4.3 4.8      5 5.2 5.8 5.006 0.3524897 50       0
```

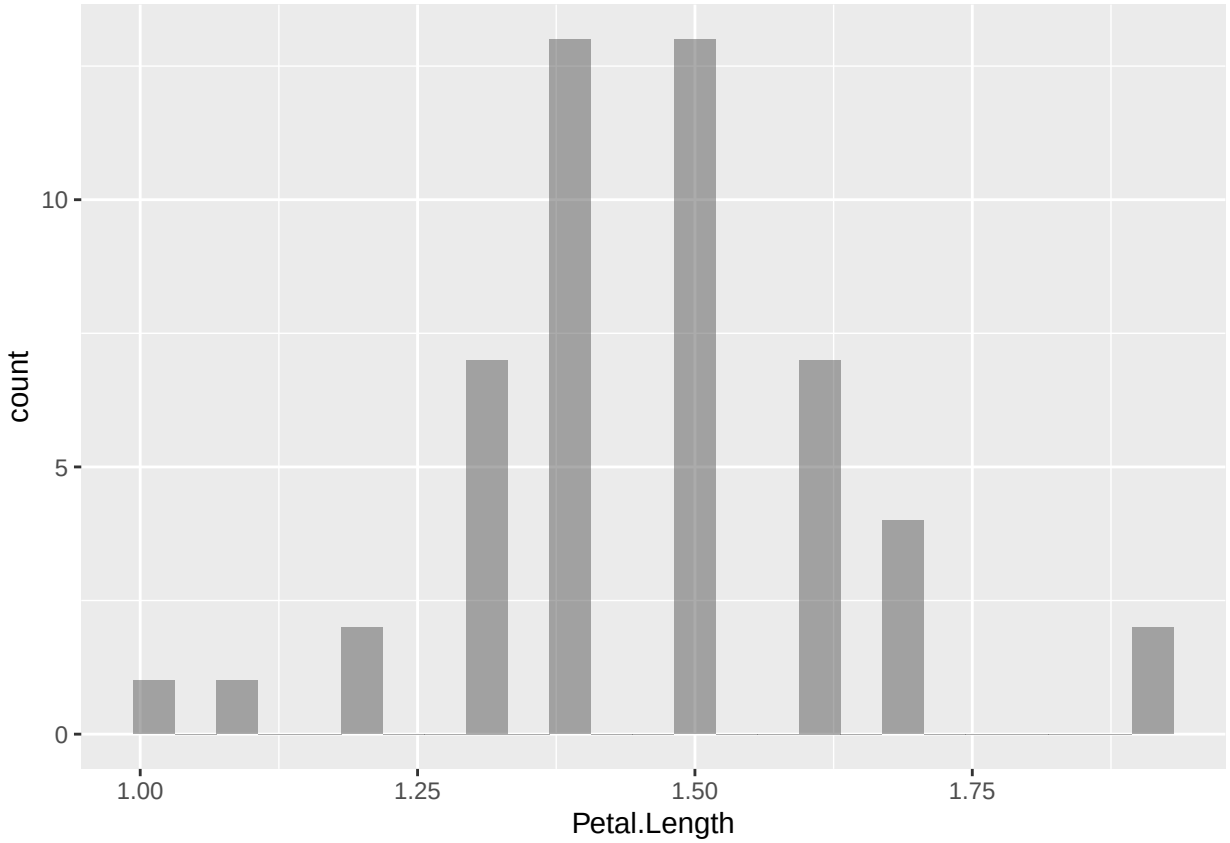
```
gf_histogram(~Sepal.Width, data = setosa)
```




```
favstats(~Sepal.Width, data = setosa)
```

```
##  min  Q1 median    Q3 max  mean        sd  n missing
##  2.3 3.2    3.4 3.675 4.4 3.428 0.3790644 50       0
```

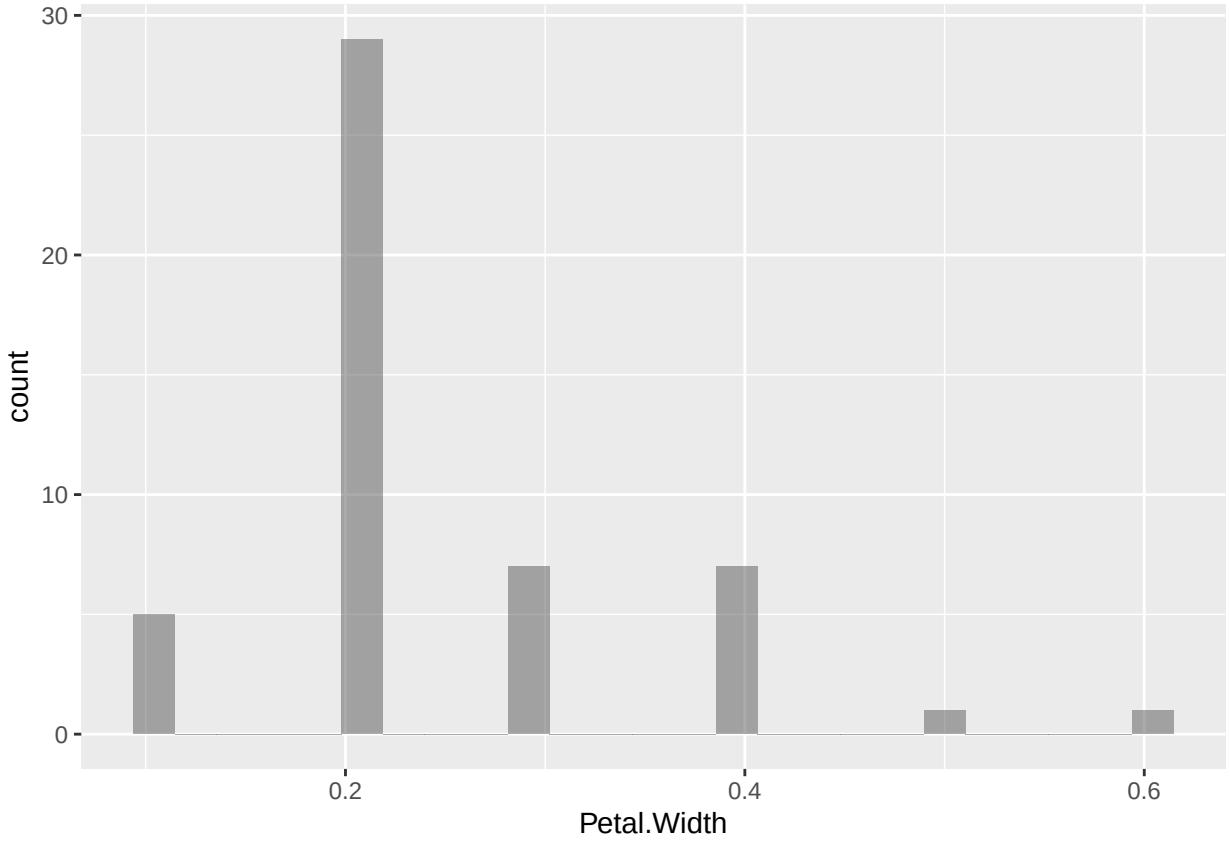
```
gf_histogram(~Petal.Length, data = setosa)
```




```
favstats(~Petal.Length, data = setosa)
```

```
##  min  Q1 median    Q3 max  mean       sd  n missing
##    1 1.4    1.5 1.575 1.9 1.462 0.173664 50       0
```

```
gf_histogram(~Petal.Width, data = setosa)
```



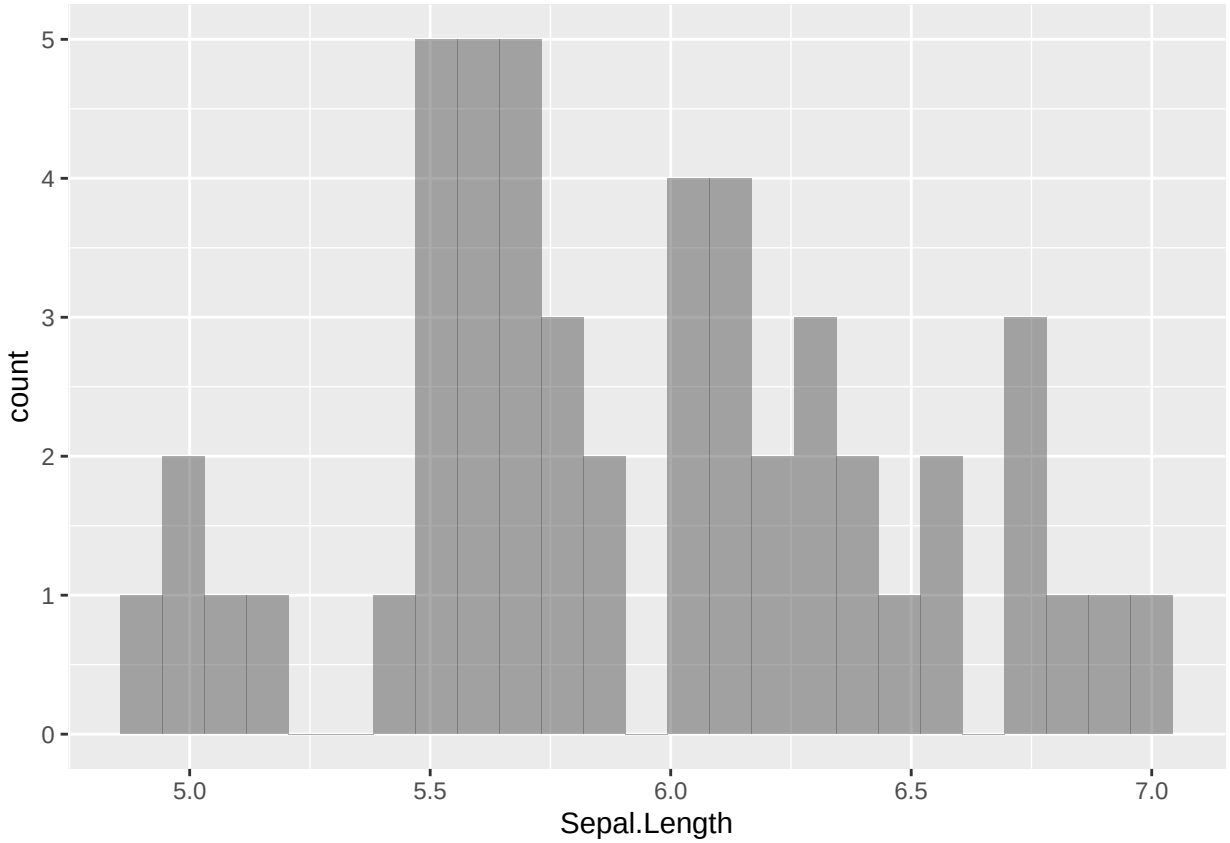

```
favstats(~Petal.Width, data = setosa)
```

```
##  min  Q1 median  Q3 max  mean        sd  n missing
##  0.1 0.2    0.2 0.3 0.6 0.246 0.1053856 50       0
```

Answer: For the setosa species, the distribution of Sepal Length seems approximately normally distributed with a mean of 5.006cm and a standard deviation of .3524897cm. The distribution of Sepal Width seems approximatley normally distributed with a mean of 3.428cm and a standard deviation of .379064cm. Petal Length also seems to be normally distributed with a mean of 1.5cm and a standard deviation of .173664cm. Lastly, Petal Width seems to have slight right skew with a median of .2cm aand a IQR of .1cm

Distributions of each variable for versicolor species

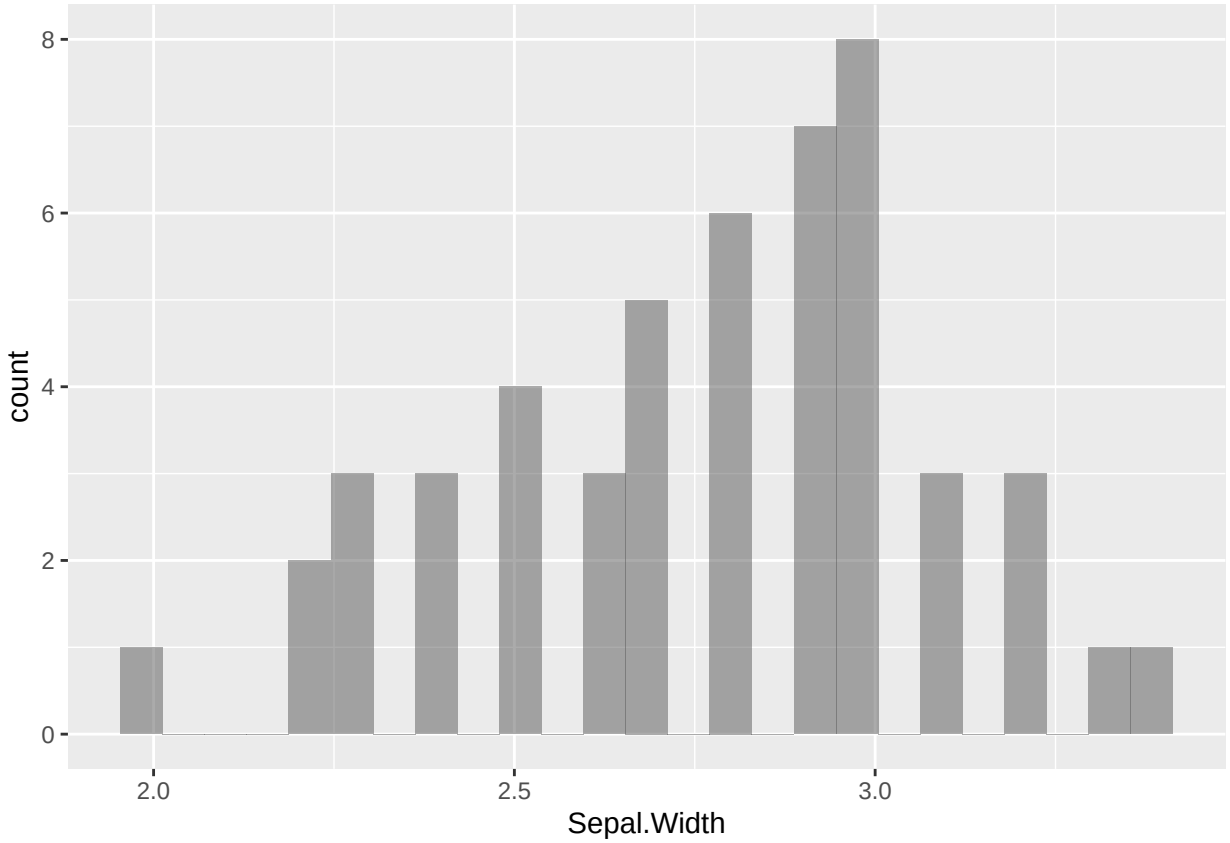
```
versicolor <- filter(iris, Species == "versicolor")
gf_histogram(~Sepal.Length, data = versicolor)
```




```
favstats(~Sepal.Length, data = versicolor)
```

```
##  min  Q1 median  Q3 max  mean        sd  n missing
##  4.9 5.6    5.9 6.3   7 5.936 0.5161711 50       0
```

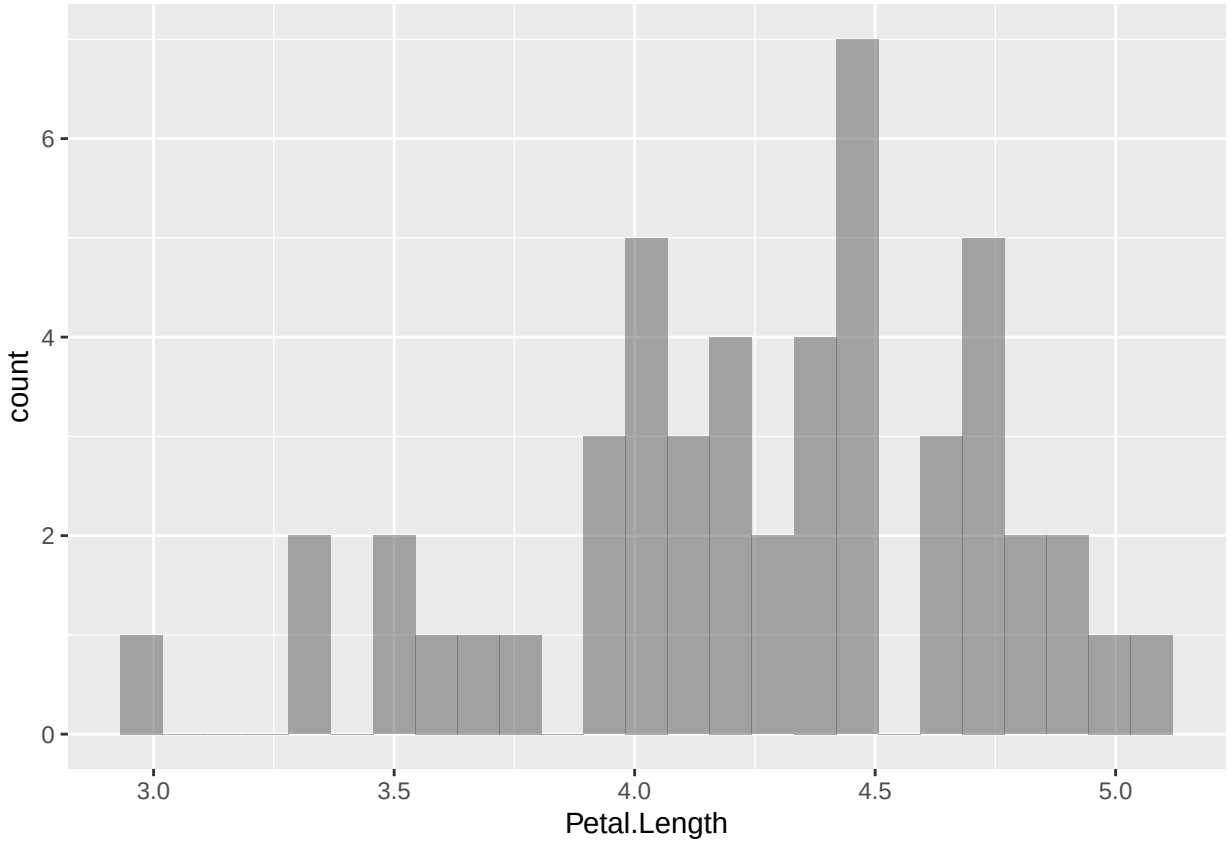
```
gf_histogram(~Sepal.Width, data = versicolor)
```




```
favstats(~Sepal.Width, data = versicolor)
```

```
##  min    Q1 median Q3 max mean        sd  n missing
##    2 2.525    2.8  3 3.4 2.77 0.3137983 50       0
```

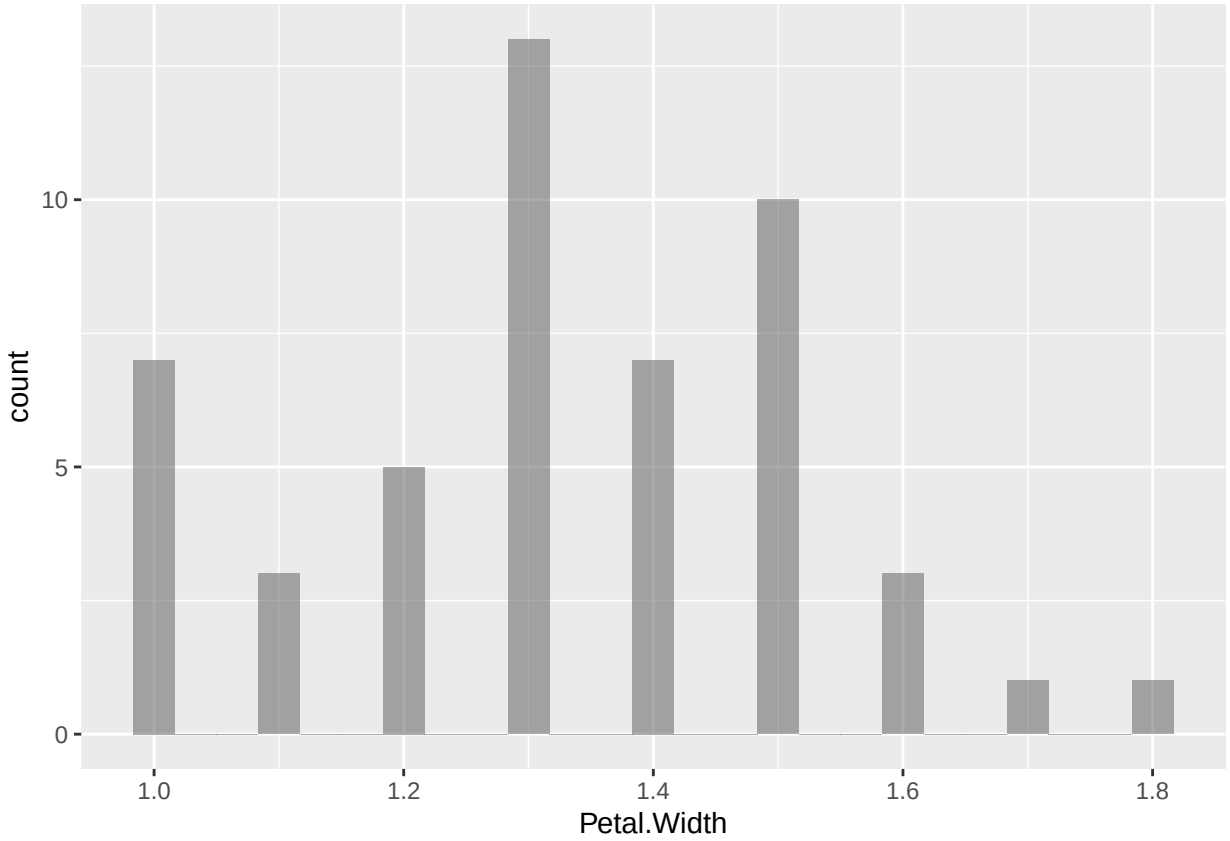
```
gf_histogram(~Petal.Length, data = versicolor)
```




```
favstats(~Petal.Length, data = versicolor)
```

```
##  min Q1 median  Q3 max mean       sd  n missing
##    3  4   4.35 4.6 5.1 4.26 0.469911 50       0
```

```
gf_histogram(~Petal.Width, data = versicolor)
```




```
favstats(~Petal.Width, data = versicolor)
```

```
##  min  Q1 median  Q3 max  mean        sd  n missing
##    1 1.2    1.3 1.5 1.8 1.326 0.1977527 50       0
```