

ggplot

2024-05-27

Import data

```
str(iris)
```

```
## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

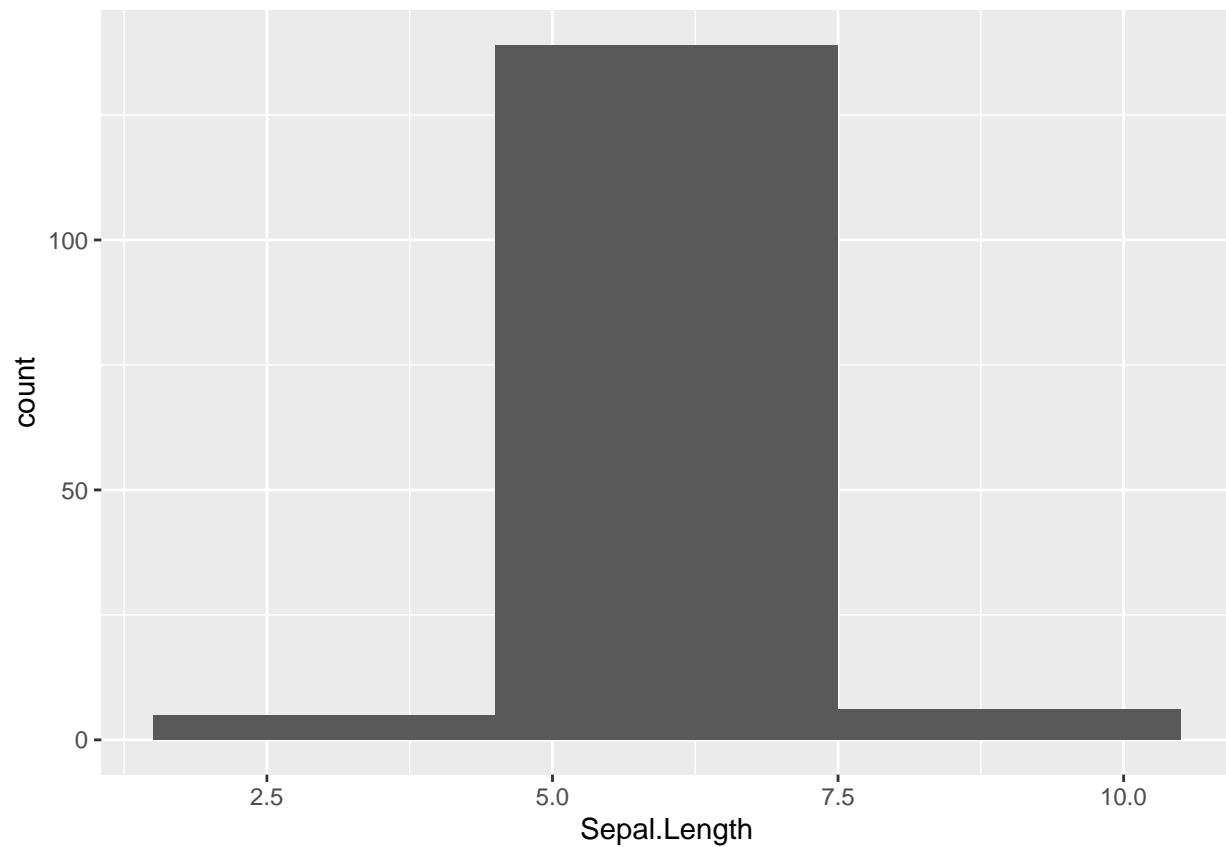
```
summary(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

```
head(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1 5.1 3.5 1.4 0.2 setosa
## 2 4.9 3.0 1.4 0.2 setosa
## 3 4.7 3.2 1.3 0.2 setosa
## 4 4.6 3.1 1.5 0.2 setosa
## 5 5.0 3.6 1.4 0.2 setosa
## 6 5.4 3.9 1.7 0.4 setosa
```

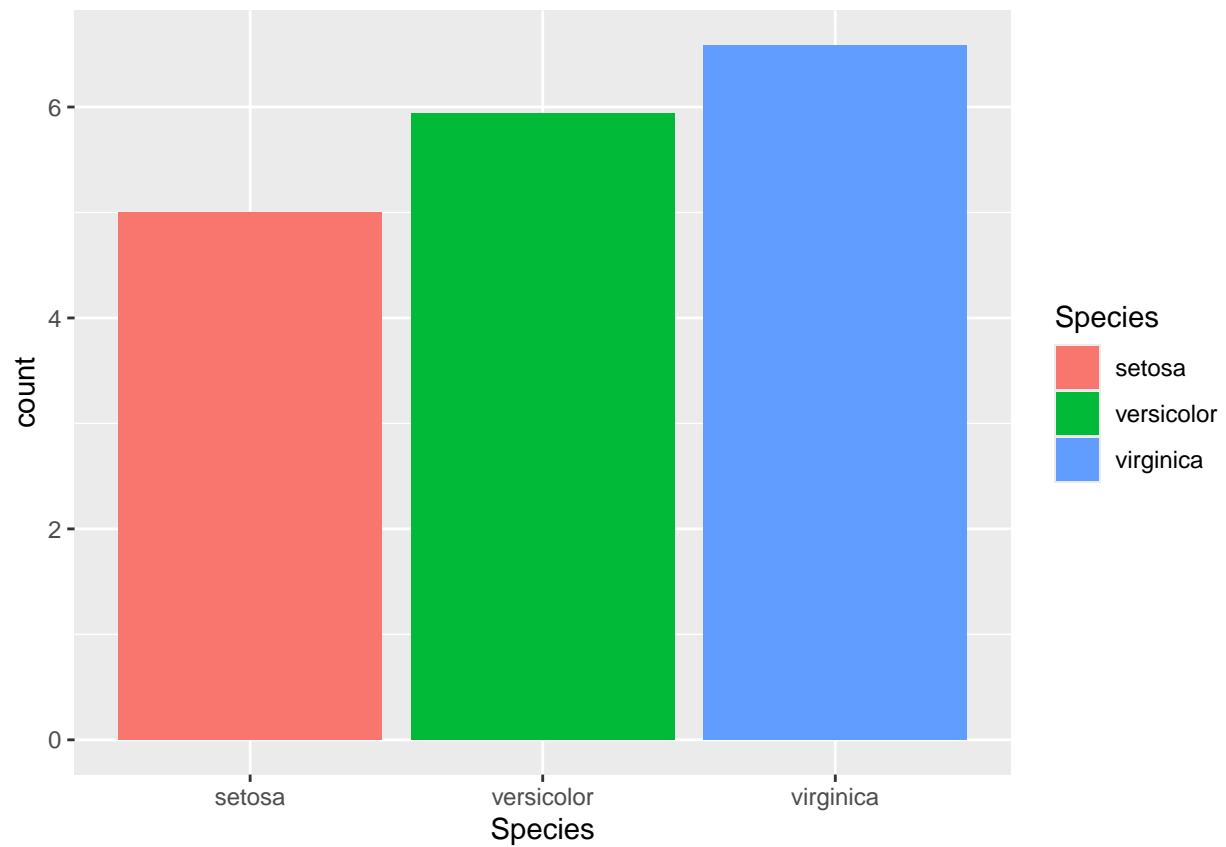
```
ggplot(iris) + geom_bar(aes(x=Sepal.Length), stat="bin", binwidth = 3)
```



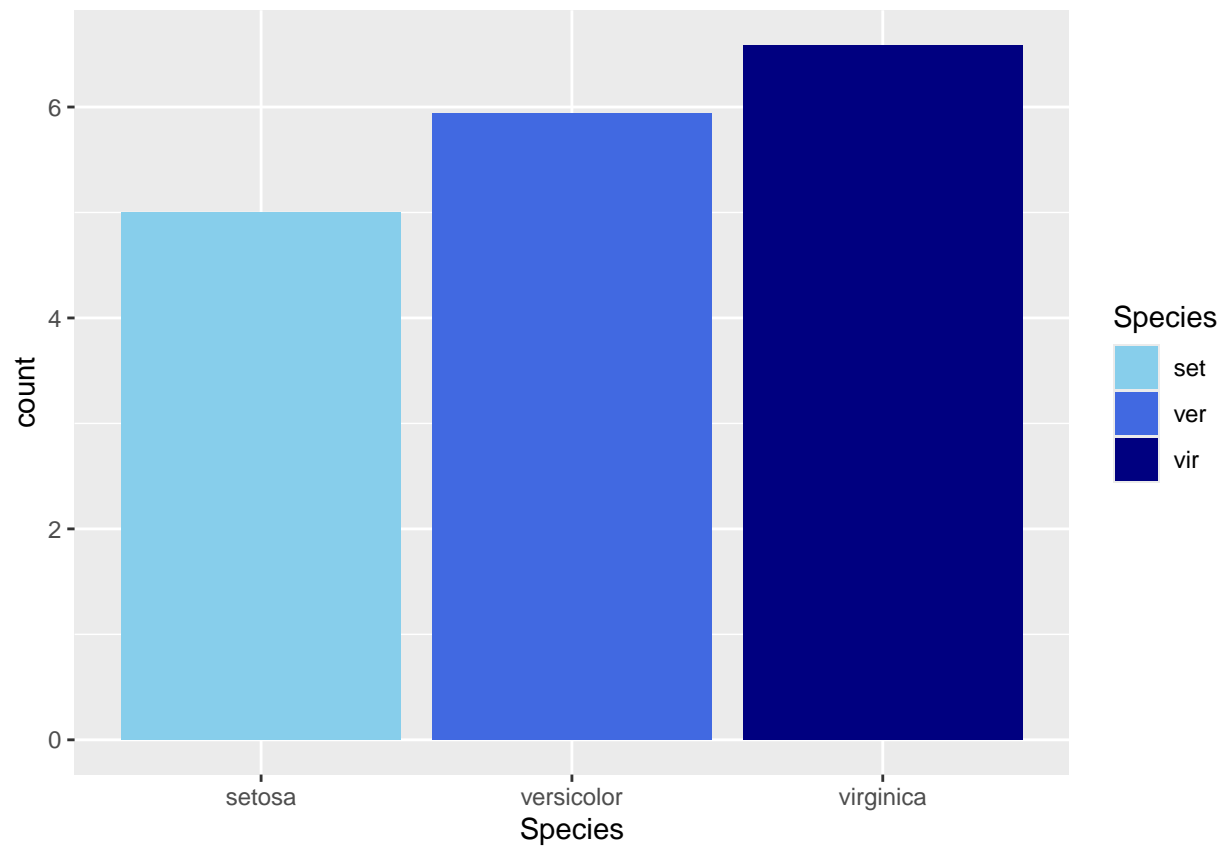
```
group_iris <- iris %>% group_by(Species) %>% dplyr::summarise(avg_sepal_length=mean(Sepal.Length))  
str(group_iris)
```

```
## tibble [3 x 2] (S3: tbl_df/tbl/data.frame)  
##   $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 2 3  
##   $ avg_sepal_length: num [1:3] 5.01 5.94 6.59
```

```
p <- ggplot(group_iris) + geom_bar(aes(x=Species, weight=avg_sepal_length, fill=Species))  
p
```



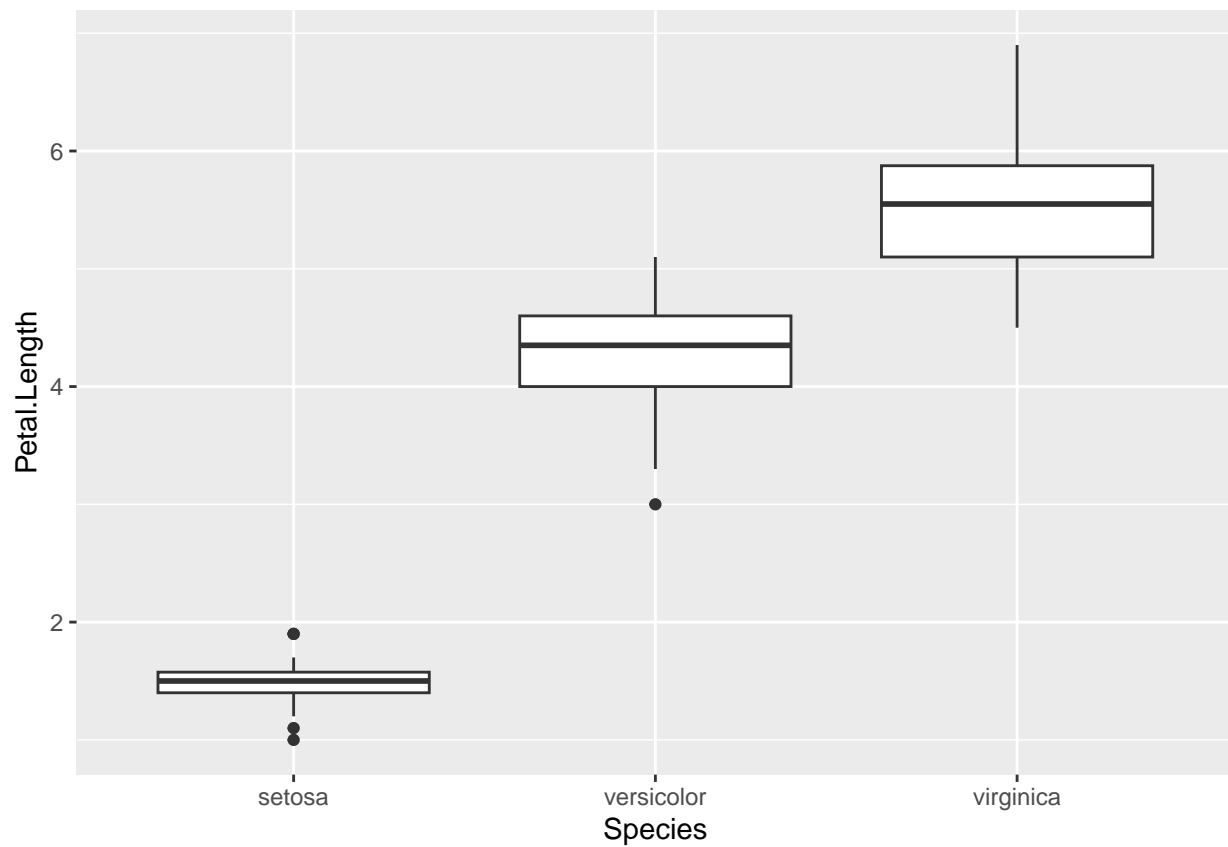
```
p <- p + scale_fill_manual(  
  values = c("skyblue", "royalblue", "navy"),  
  limits = c("setosa", "versicolor", "virginica"),  
  breaks = c("setosa", "versicolor", "virginica"),  
  name = "Species",  
  labels = c("set", "ver", "vir")  
)  
p
```



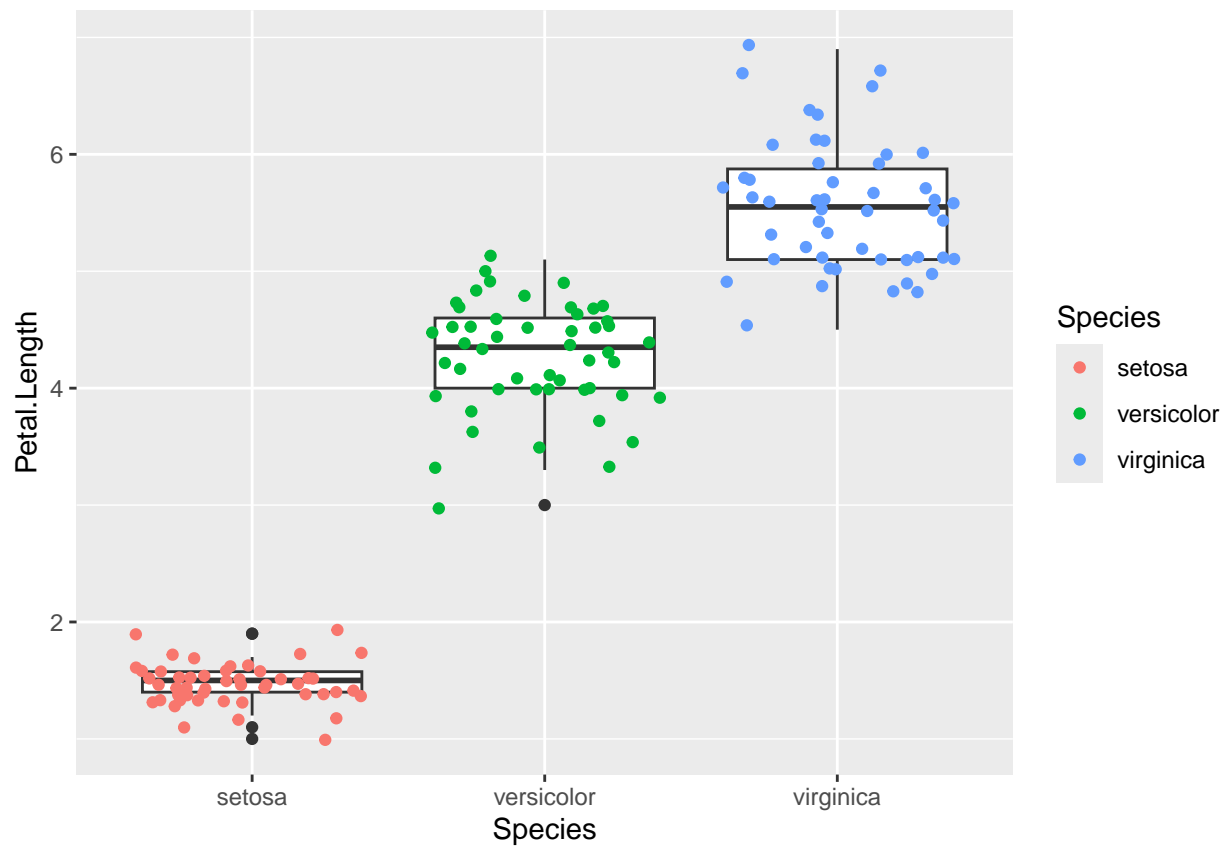
```
f.scatter <- ggplot(iris) + geom_point(aes(x=Sepal.Length, y=Sepal.Width, color=Species))  
f.scatter
```



```
# Boxplot to explore numeric variable
p.box <- ggplot(iris) + geom_boxplot(aes(x=Species, y=Petal.Length))
p.box
```



```
# One way we can extend this plot is adding a layer of individual points on top of it
p.box.jitter <- p.box + geom_jitter(aes(x=Species, y=Petal.Length, color=Species))
p.box.jitter
```



```
# A final plot useful for looking at univariate relations is the kdeplot,
p.density <- ggplot(iris) + geom_density(aes(x=Petal.Length, colour=Species))
p.density
```

