

$$(X, O) = e^{-\frac{x^2}{2\sigma^2}}$$

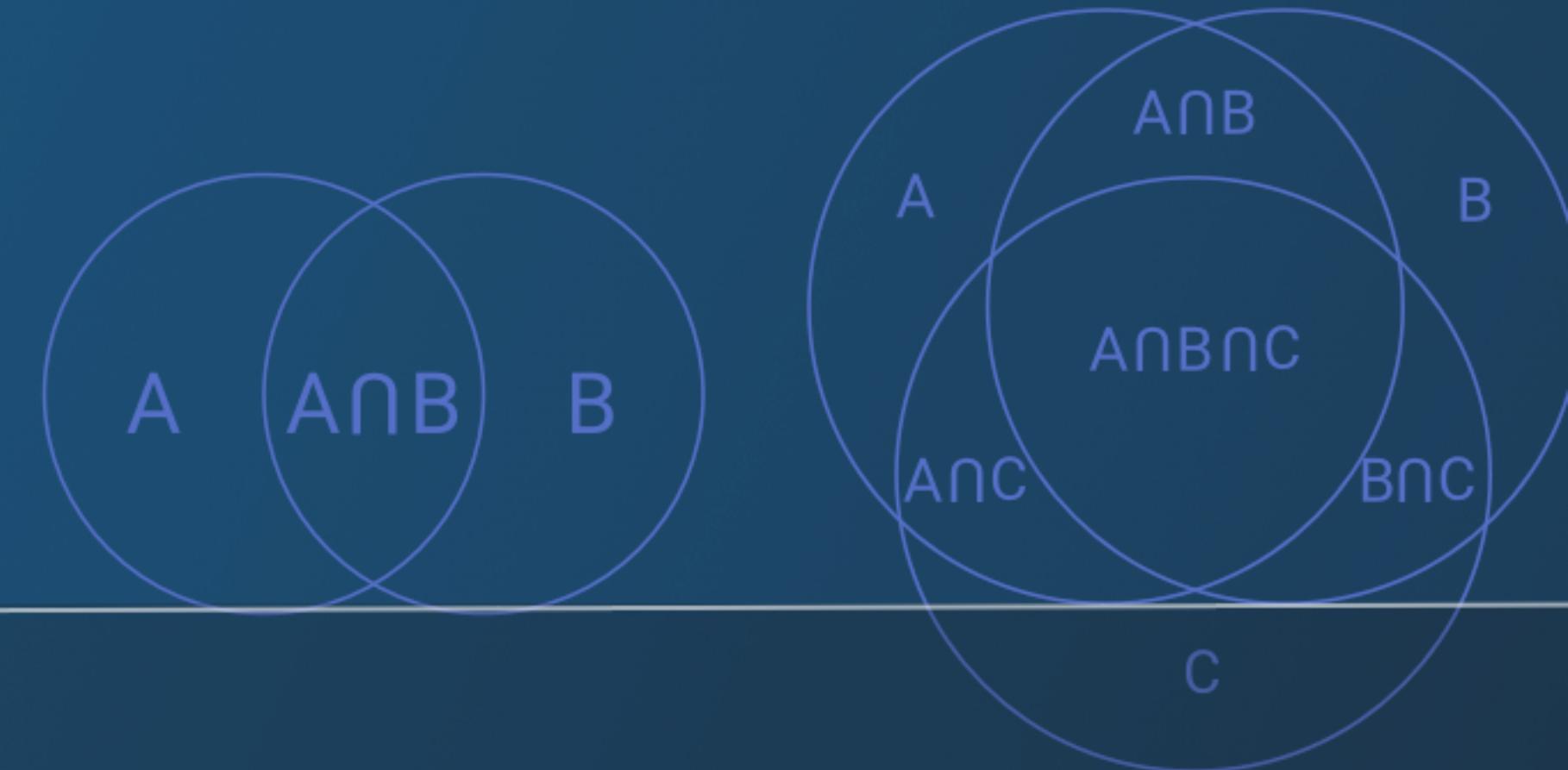
$$x(X, O) = -\frac{x}{\sigma^2} G(X, O) = -\frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

$$xx(X, O) = \frac{x^2 - \sigma^2}{\sigma^4} G(X, O) = \frac{x^2 - \sigma^2}{\sigma^4} e^{-\frac{x^2}{2\sigma^2}}$$

$$xxx(X, O) = -\frac{x^3 - x\sigma^2}{\sigma^6} G(X, O) = -\frac{x^3 - x\sigma^2}{\sigma^6} e^{-\frac{x^2}{2\sigma^2}}$$

$$= x \ln(x + \sqrt{1 + x^2})$$

# Julia 程式語言學習馬拉松



## Day 21



cupay

陪跑專家 : Andy Tu

# Julia 基礎圖表繪製

## Gadfly 一維資料圖表





# 重要知識點



- 認識一維資料圖表
- 了解不同的一維資料圖表的繪製方式
- 了解不同的一維資料圖表的適用情境



# DataFrame：建立



- 在資料分析中，資料視覺化是重要的一環。一維資料圖表是指將資料的其中一個維度以圖表的方式做呈現。
- 一維資料圖表可以讓我們了解資料中單一欄位的資料特性。
- 常見的一維資料圖表：
  - 直方圖
  - 條狀圖
  - 線圖
  - 盒狀圖
  - Violin plot



# 為什麼我們需要一維資料圖表？



- 往往在看資料時，人類對於數值的敏感度不如圖表。
- 當資料在一百筆以內，尚可以觀看的方式理解資料，但在這個數字以上，是人腦難以處理的。
- 在看資料時，人類對於類別出現的頻率也較不敏感。



# Iris 資料集



- 這邊我們以 iris 資料集作為示範。
- 在 Julia 中可以從 RDatasets 套件讀取資料集。

```
using RDatasets  
iris = dataset("datasets", "iris")
```



# Gadfly



- Gadfly 是一套現代的繪圖套件，有簡單易懂的語法，以及簡潔的圖表呈現。
- 它大量地參考了 R 語言的 ggplot2 引擎，以及 The Grammar of Graphics 一書。
- 官方網站：<http://gadflyjl.org/stable/>

using Gadfly



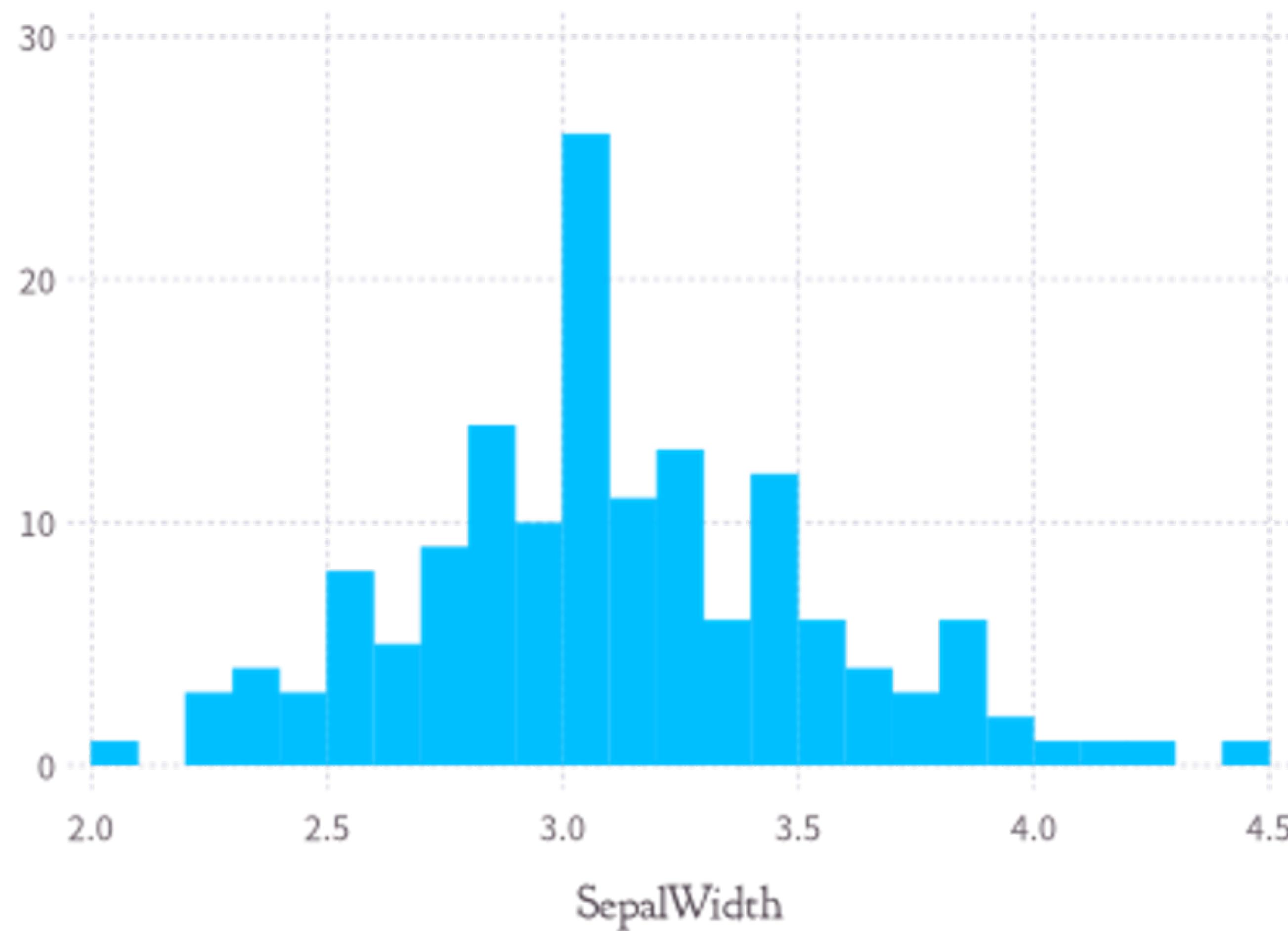
# 直方圖

- 直方圖 (histogram) 是經常使用在連續型變數的一維資料圖表，可以用來觀察資料出現的頻率。
- 直方圖也可以用來比較不同區間的資料出現的頻率。
- 使用上需要給定 DataFrame 並指定要繪製的資料欄位給 x，最後需要加上 Geom.histogram。

```
plot(iris, x="SepalWidth", Geom.histogram)
```



# 直方圖





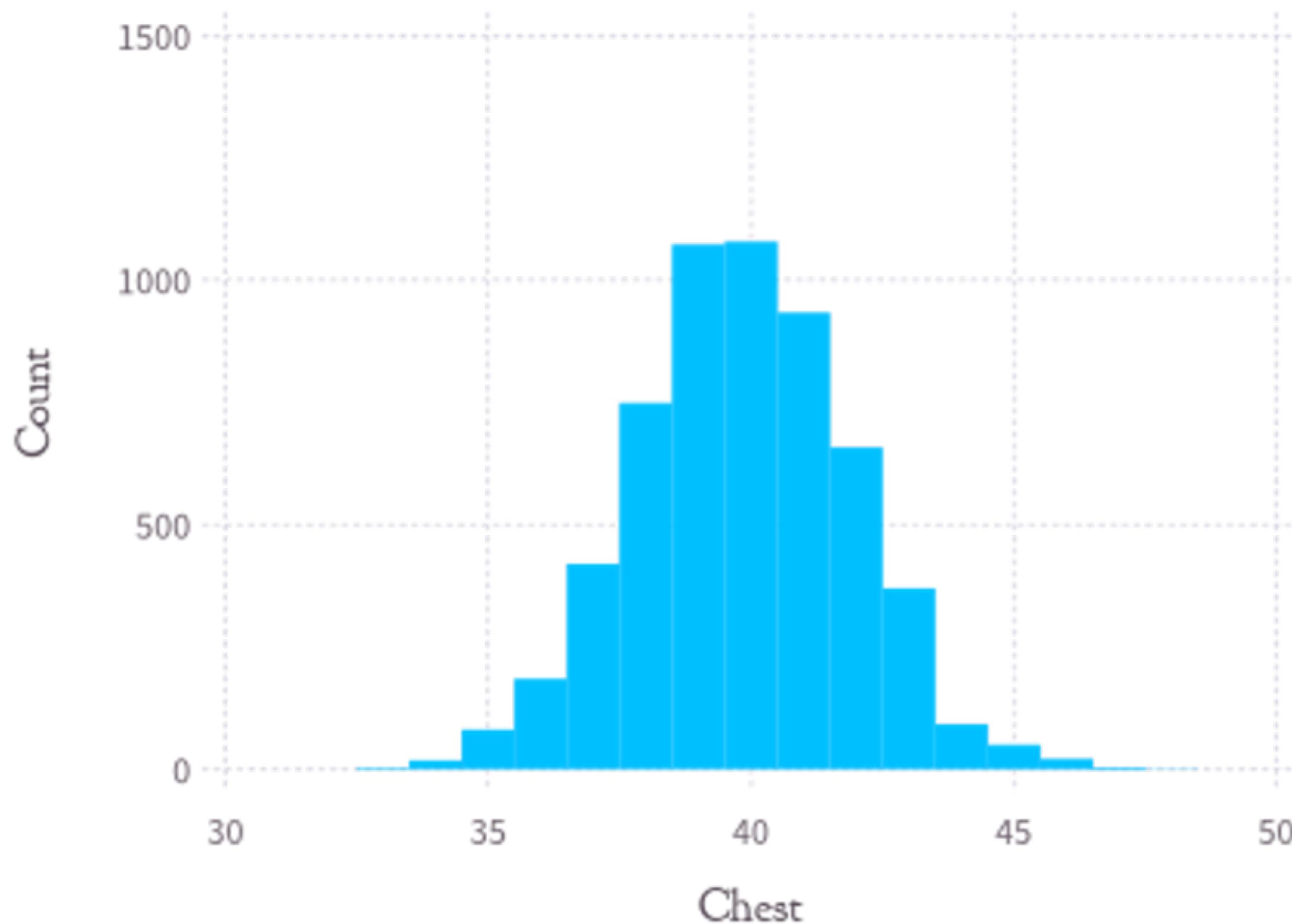
# 條狀圖

- 條狀圖 (bar chart) 也是一種觀察資料出現頻率的方式，不過與直方圖不同的是它是應用在離散型變數上。
- 條狀圖有與直方圖相似的應用場景與特性。
- 使用上需要給定 DataFrame 並指定要繪製的離散型資料欄位給 x，以及相對應的計數給 y，最後需要加上 Geom.bar。

```
chest = dataset("HistData", "ChestSizes")
plot(chest, x="Chest", y="Count",
Geom.bar)
```



# 條狀圖





# Gadfly



- 嘗試過兩種圖表後，應該大致了解 Gadfly 的語法規則。
- 繪圖只需要 plot 函式，並給定 DataFrame。
- 指定資料欄位到相對應的軸上，例如，`x="Chest", y="Count"`。
- 指定要繪製的表格類型，例如，`Geom.bar`。
- 以上的參數順序沒有關係。

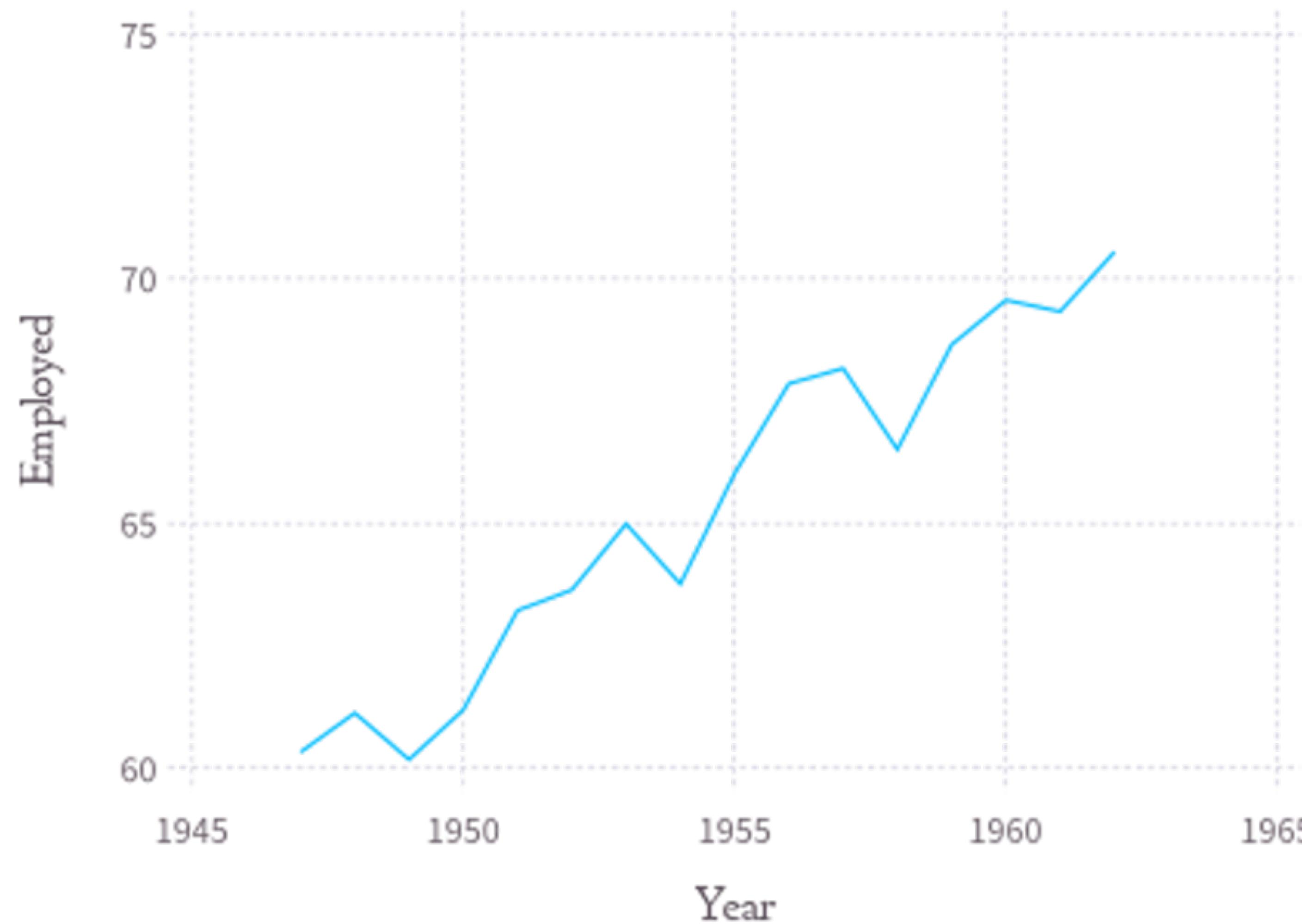


# 線圖

- 線圖 (line chart) 是個可以展現資料趨勢的圖表。
- 接下來，我們使用一個總體經濟的資料集來看看趨勢，我們將年份放在 x 軸，僱員人數放在 y 軸，再加上 Geom.line 就可以畫出線圖了。
- 可以搭配 Geom.point 試著畫畫看有什麼效果。

```
longley = dataset("datasets", "longley")
plot(longley, x="Year", y="Employed",
Geom.line)
```

# 線圖



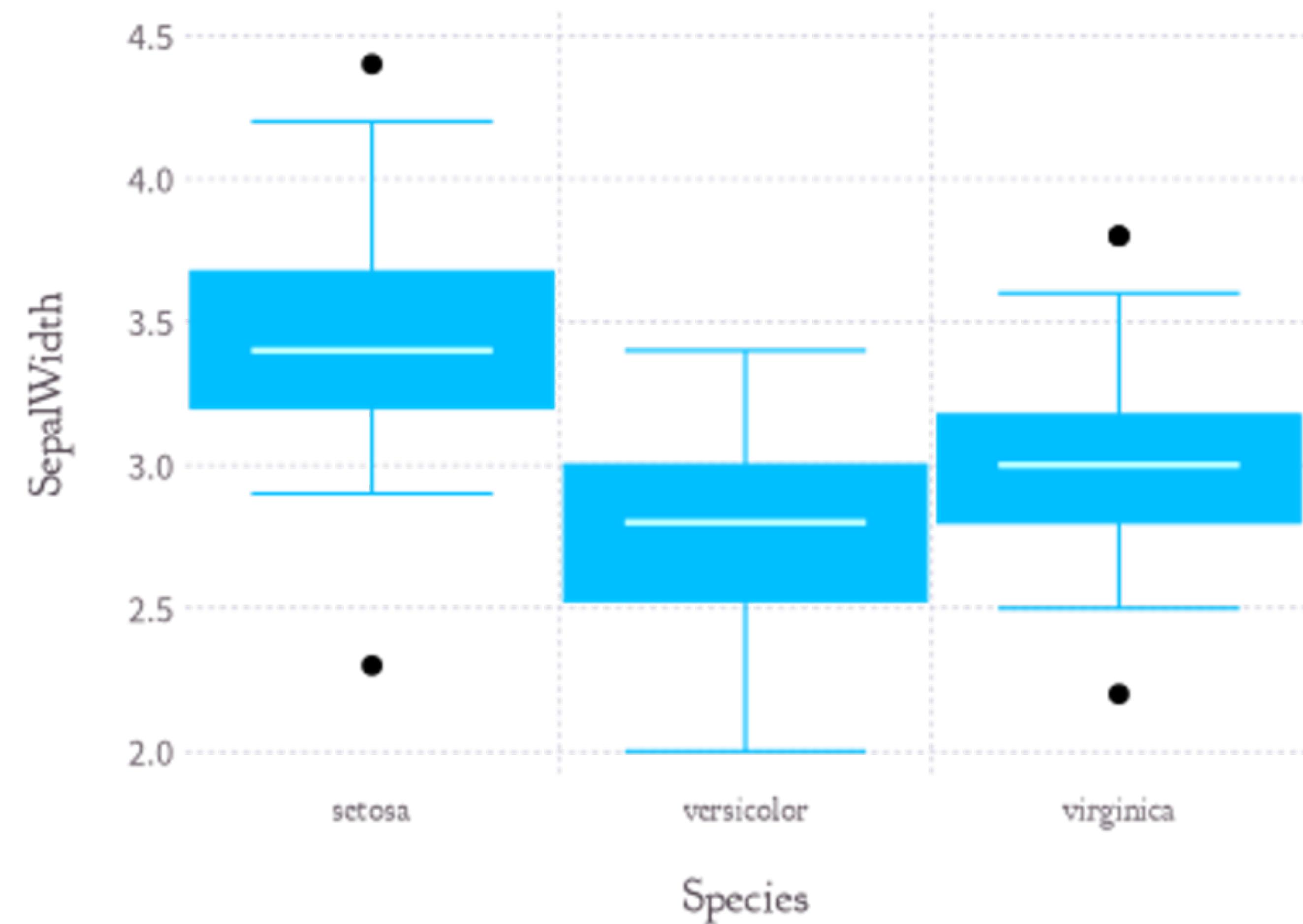


# 盒狀圖

- 盒狀圖 (boxplot) 可以將不同類別分別畫出它們資料的分佈狀況，他可以顯示出資料的最大值、最小值、中位數、及上下四分位數。
- 盒狀圖可以用來比較類別之間資料的分佈狀態。
- x 軸會是離散型變數，y 軸則是連續型變數。

```
plot(iris, x="Species", y="SepalWidth",
Geom.boxplot)
```

# 盒狀圖





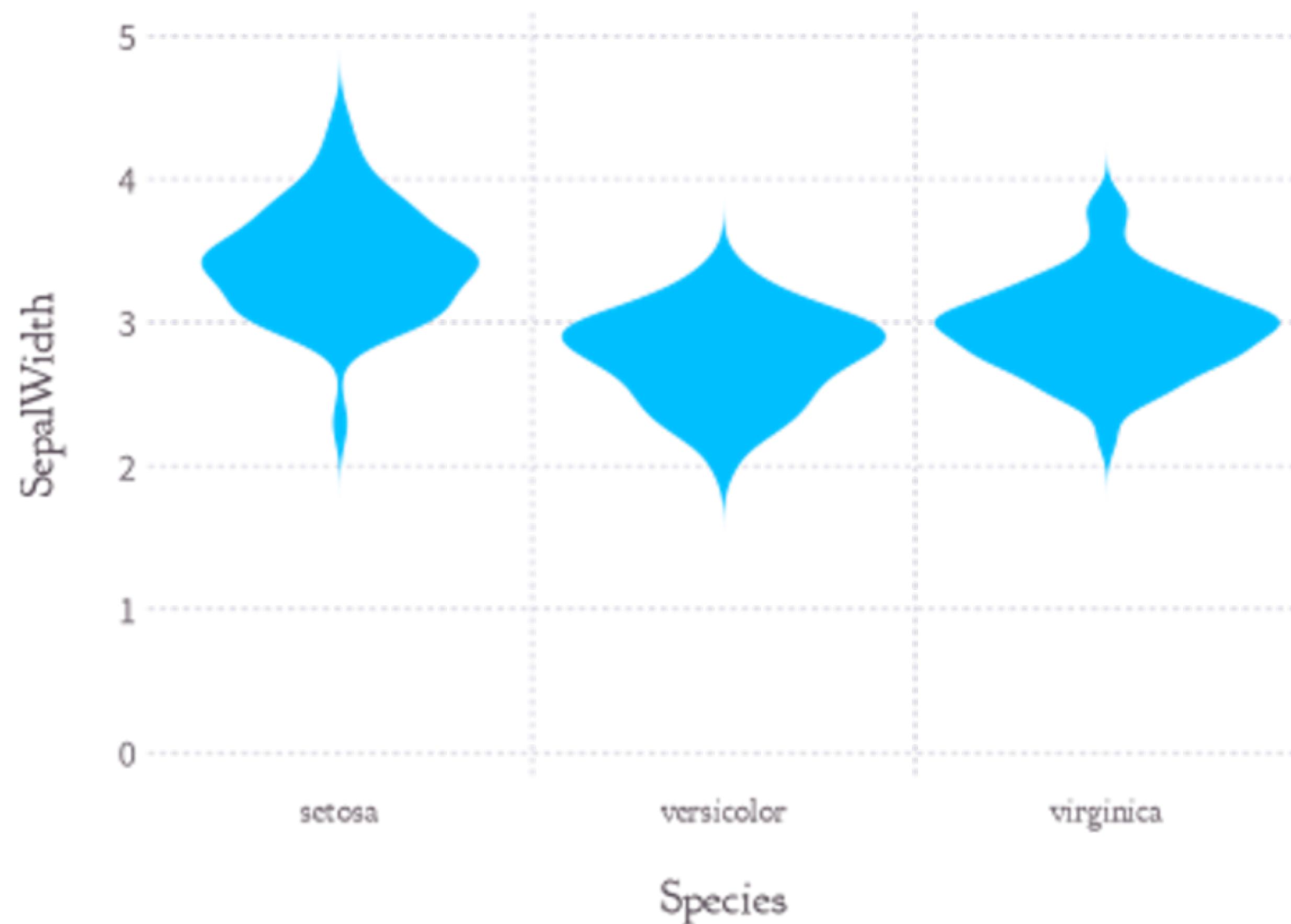
# Violin Plot

- Violin plot 也是將不同的類別分別畫出它們資料分佈的情況，與盒狀圖不同的是，violin plot 會去估計出資料分佈的密度。
- Violin plot 可以用來比較類別之間資料的分佈狀態。
- x 軸會是離散型變數，y 軸則是連續型變數。

```
plot(iris, x="Species", y="SepalWidth",
Geom.violin)
```



# Violin Plot



# 知識點 回顧

- 認識一維資料圖表，及其繪製方式、適用情境
  - 直方圖
  - 條狀圖
  - 線圖
  - 盒狀圖
  - Violin plot



## 推薦閱讀

- [Gadfly.jl 官方網頁](#)
- [RDatasets.jl 程式庫](#)





解題時間

請跳出 PDF 至官網 Sample Code  
& 作業開始解題