In [2]:
```
using RDatasets;
data = dataset("car","Davis");
```

1b

In [5]:
```
weight = data[:, 2]
height = data[:, 3]
n,_ = size(data)
mu_weight = sum(weight)/n
mu_height = sum(height)/n
var_weight = sum((weight - mu_weight).^2)/n
var_height = sum((height - mu_height).^2)/n

println("The mean of Weight is ", mu_weight)
println("The variance of Weight is ", round(var_weight,2))
println("The mean of height is ", mu_height)
println("The variance of height is ", round(var_height, 2))
```

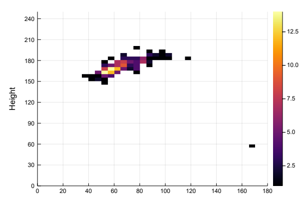
```
The mean of Weight is 65.8
The variance of Weight is 226.72
The mean of height is 170.02
The variance of height is 143.47
```

1c

This looks like a MV Gaussian where Height & Weight have more middle values & less left & right values.

In [14]:
```
using Plots;
gr()
histogram2d(data[:Weight],data[:Height],nbins=40,xlabel="Weight",ylabel="Height",
```

Out[14]:



1d

In [22]:
```
mv_data = data[:,2:3]
mu_mv = [mu_weight mu_height]
cov_height_weight = (sum((weight - mu_weight).*(height - mu_height))) / n
cov_matrix = zeros(Float64, 2, 2)
cov_matrix[1,1] = var_weight
cov_matrix[2,2] = var_height
cov_matrix[1,2] = cov_matrix[2,1] = cov_height_weight
cov_matrix
```

Out[22]:
```
2×2 Array{Float64,2}:
 226.72    34.204
  34.204  143.47
```

1e

Yes it appears to be a good fit for the data

In [23]:
```
using Distributions;
d = MvNormal([65.8;170.02], [226.72 34.204; 34.204 143.47])
```

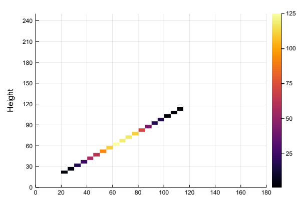
Out[23]:
```
FullNormal(
dim: 2
μ: [65.8, 170.02]
Σ: [226.72 34.204; 34.204 143.47]
)
```

In [33]:
```
sample = rand(d,1000)
```

Out[33]:
```
2×1000 Array{Float64,2}:
  69.9483    76.1801    42.0015    59.5669  …   63.8352    63.4679    66.041
 157.546    178.355    174.411    173.051      183.658    161.866    182.174
```

In [34]:
```
histogram2d(sample[1,:],sample[1,:],nbins=40,xlabel="Weight",ylabel="Height", xax
```

Out[34]:



1f No they're not dependent

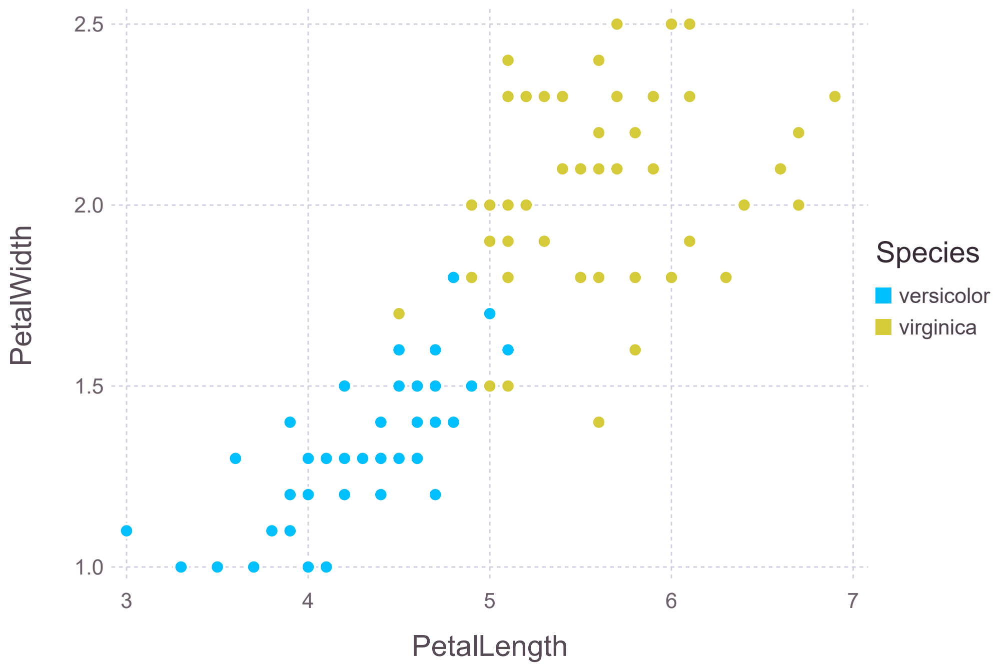
2a

In [37]:
```
using RDatasets;
iris = dataset("datasets", "iris");
groups = groupby(iris,:Species);
new_iris = vcat(groups[2],groups[3]);
```

2b

In [38]:
```
using Gadfly;
Gadfly.plot(new_iris, x=:PetalLength, y=:PetalWidth,color=:Species,Geom.point)
```

Out[38]:



2h

2. d. Let Versicolor = 1 and ~~Pet~~ Virginica = 0

Let

Petal Length = x_1 and Petal Width = x_2

$$f(y=1, x) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

where σ is sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$

Here y is binary variable $\in \{0, 1\}$

and x_1, x_2 are continuous variable $\in R$

e. Conditional Likelihood

$$= \prod_i f(y_i | x_i ; \beta)$$

where $i : \{1, 2 \cdots n\}$; $x_i = [1 \; x_{1i} \; x_{2i}]$; $\beta = [1 \; \beta_0 \; \beta_1]$

n is the sample size

f. log conditional Likelihood

$$= \sum_i \log f(y_i | x_i ; \beta)$$

splitting sum where $y_i = 1$ and $y_i = 0$

Let $p_i = p(y_i = 1 | x_i ; \beta)$

$\therefore$ log conditional likelihood $= \sum_{i, y_i = 1} \log p_i + \sum_{i, y_i = 0} \underbrace{\log(1-p_i)}_{\Downarrow}$

$\log(f(y_i = 0 | x_i, \beta))$

2.e $l = \sum_{i, y_i=1} \log p_i + \sum_{i, y_i=0} \log(1-p_i)$

$$\vec{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2]$$

Partial Derivative of l w.r.t β_j

$$(A) \text{—} \quad \frac{\partial l}{\partial \beta_j} = \sum_{i, y_i=1} \frac{1}{p_i}\underbrace{\left(\frac{\partial p_i}{\partial \beta_j}\right)}_{1} + \sum_{i, y_i=0} \frac{1}{(1-p_i)}\underbrace{\left(\frac{-\partial(p_i)}{\partial \beta_j}\right)}_{2}$$

Finding 1 and 2

$$\text{Let} \quad e = \exp\left[-\sum_{j=0}^{d} \beta_j x_j\right] \qquad p_i = \frac{1}{1+\exp^{-[\Sigma \beta_i x_j]}}$$

$$\therefore p = \frac{1}{1+e}$$

$$1-p = \frac{e}{1+e}$$

$$\therefore \frac{\partial p}{\partial \beta_j} = \frac{-1}{(1+e)^2}\left(\frac{\partial e}{\partial \beta_j}\right)$$

$$= \frac{-1}{(1+e)^2} \times e \times \frac{\partial(-\Sigma \beta_j x_j)}{\partial \beta_j}$$

$$= \frac{+e x_j}{(1+e)^2} = p(1-p)x_j$$

Substituting it in A

$$\frac{\partial l}{\partial \beta_j} = \sum_{i, y=1} \frac{1}{p_i}\left(p_i(1-p_i)\, x_{ij}\right) + \sum_{i, y_i=0} \frac{-1}{(1-p_i)}\left(p_i(1-p_i)\, x_{ij}\right)$$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i, y_i=1} (1-p_i)\, x_{ij} - \sum_{i, y=0} p_i x_{ij}$$

Collecting both terms

$$\frac{\partial l}{\partial \beta_j} = \sum_i (y_i - p_i)\, x_{ij}$$

Thus

$$\frac{\partial l}{\partial \beta_0} = \sum_i (y_i - p_i) \qquad (\text{Since } x_0 = 1)$$

where $i : \{1, 2 \ldots n\}$ $n \to$ sample size

$$\frac{\partial l}{\partial \beta_1} = \sum_i (y_i - p_i)\, x_{1i} \qquad x_1 = \text{Petal Length}$$

$$\frac{\partial l}{\partial \beta_2} = \sum_i (y_i - p_i)\, x_{2i} \qquad x_2 = \text{Petal Width}$$