Certification Program on Business Analytics

Assignment No:4

Date: 5th March to 10th March-2018

Submitted By:
Mallesham Yamulla

Topics

1. Read and Understand the Two Papers/articles titled

   (i) Fitting Distribution to Hurricane Data
   (ii) Fitting Distributions with R

   Make a brief Summary on the main theme of each paper.

2. Identify a suitable dataset (Health Statistics or some interesting data set). Identify the distribution for each of the characteristic in Health Statistics Data. Set limits known to you as limits (For example Systolic : 120-130 and Diastolic: 75-85 etc and compute the abnormal proportions/probability in each case.

3. Complete the distribution identification of the remaining characteristics not discussed in the class using either MINITAB or using QQ Plot of R

4. Select 20 questions from the following questions of chapters 3 & 4 of CBS by Aczel

Fitting distribution with R

Fitting distributions consists in finding a mathematical function which represents in a good way a statistical variable. A statistician often is facing with this problem: he has some observations of a quantitative character x_1, x_2,... x_n and he wishes to test if those observations, being a sample of an unknown population, belong from a population with a pdf (probability density function) $f(x,\theta)$, where θ is a vector of parameters to estimate with available data.

We can identify 4 steps in fitting distributions:

1) Model/function choice: hypothesise families of distributions;

2) Estimate parameters;

3) Evaluate quality of fit

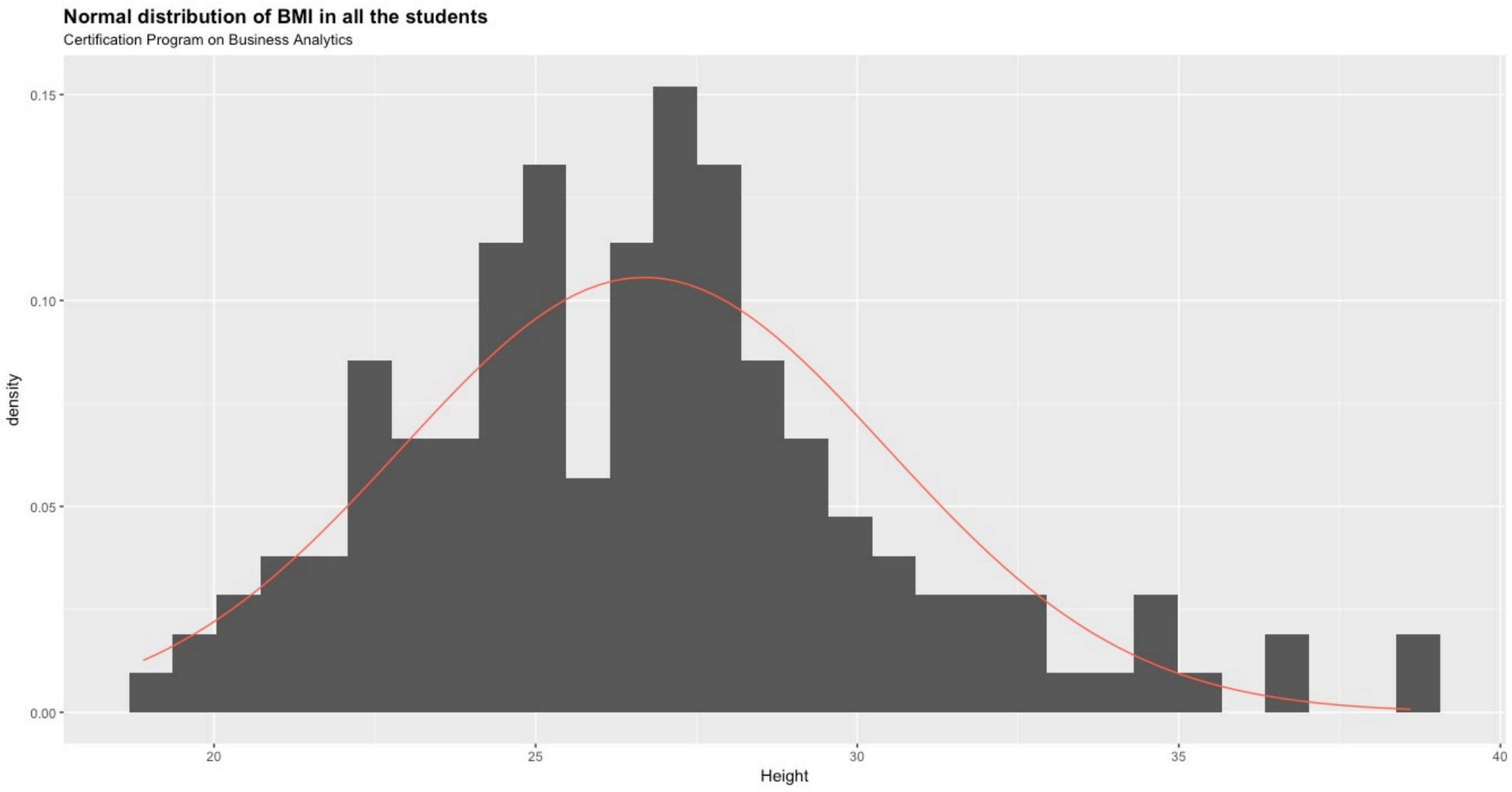
4) Goodness of fit statistical tests.

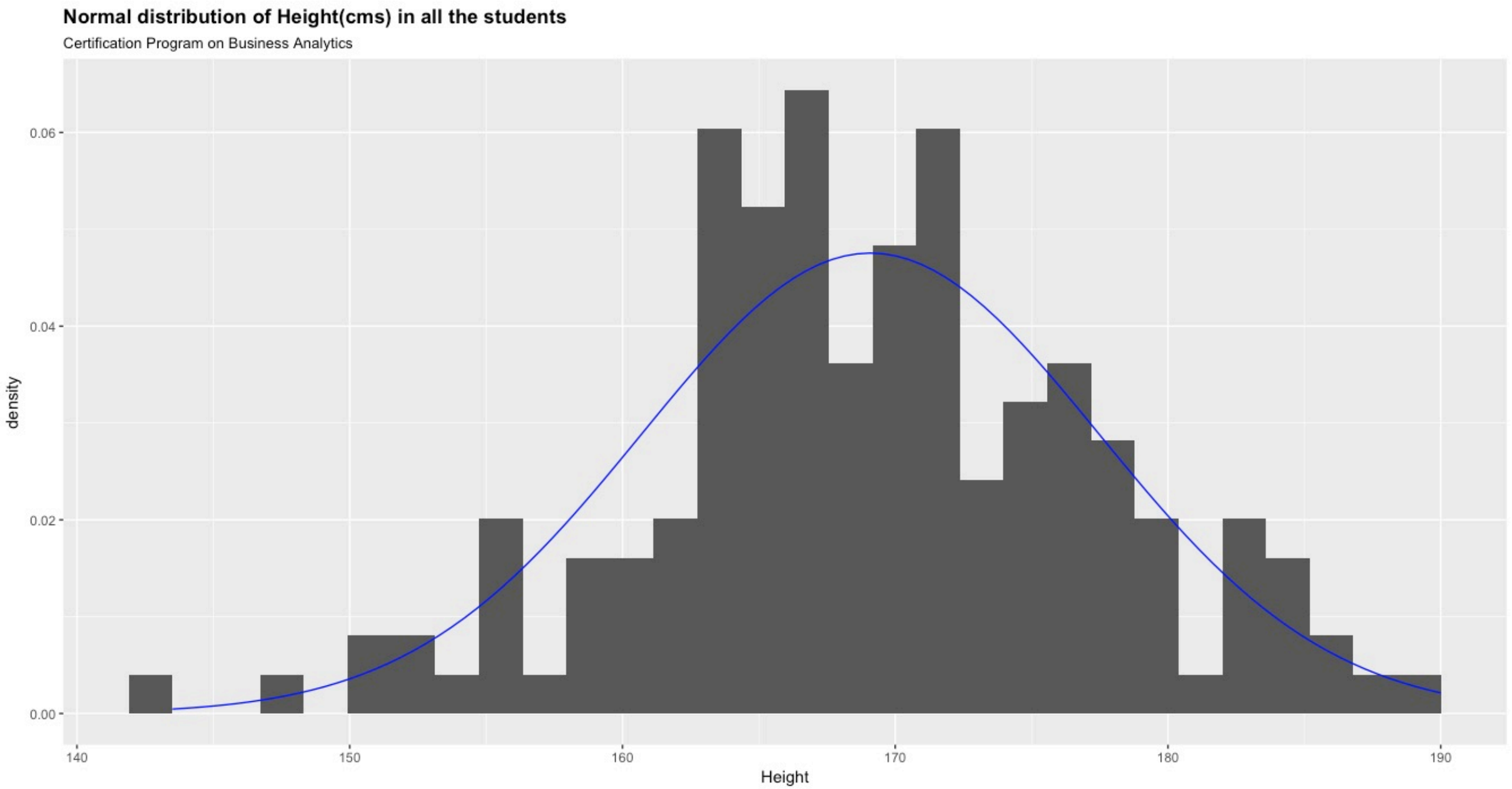
We can use probability density functions to fit the model, this can be achieved by carrying the below procedures sequentially

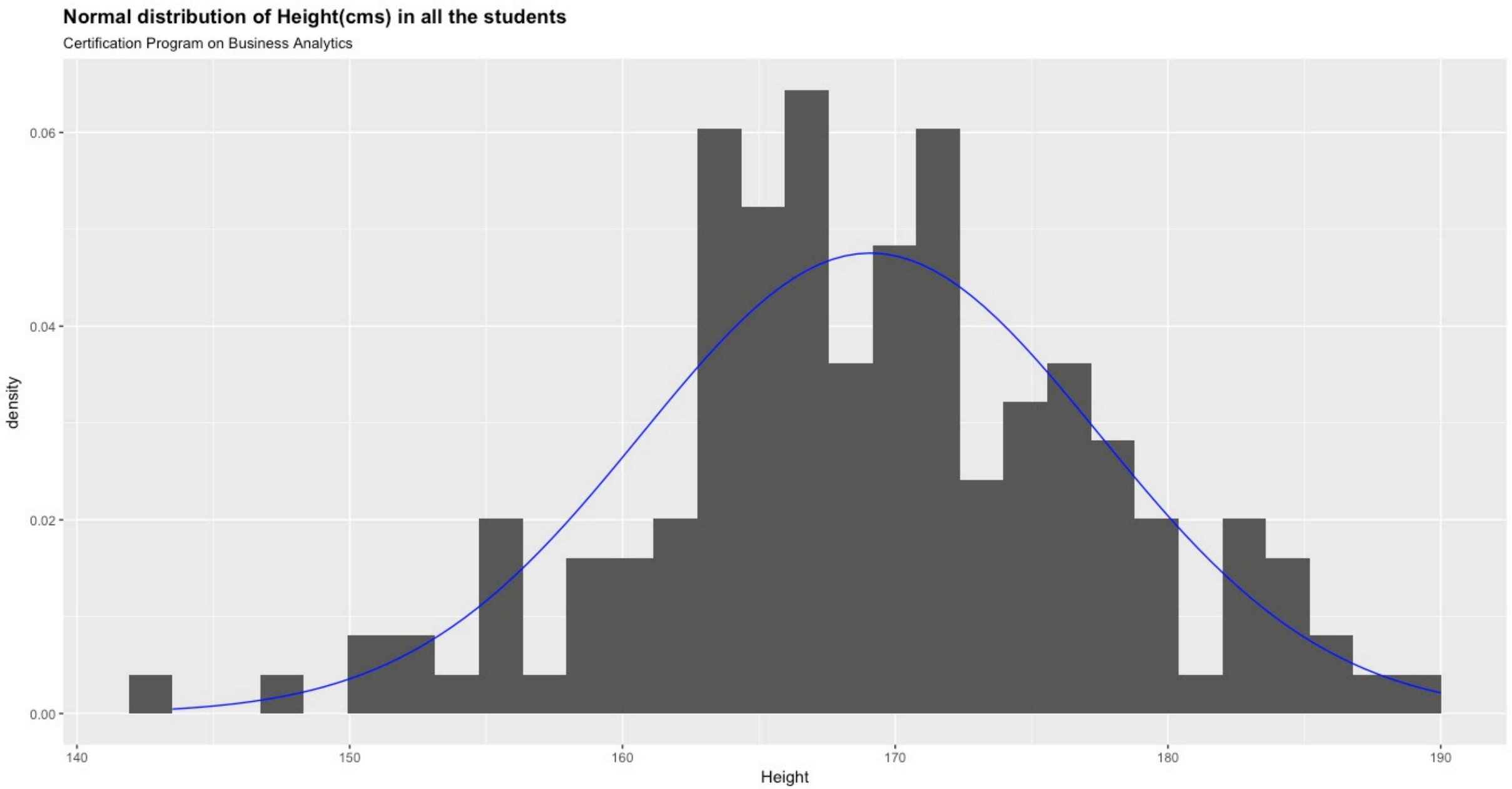
1. Exploratory data analysis
2. descriptive statistics
3. Graphical visualisations such as(Histograms, line/density plots etc.etc.)

I have tried to implement the above procedure to the health stats data using R and presented in the next slides.

2 and 3rd Questions-Answers

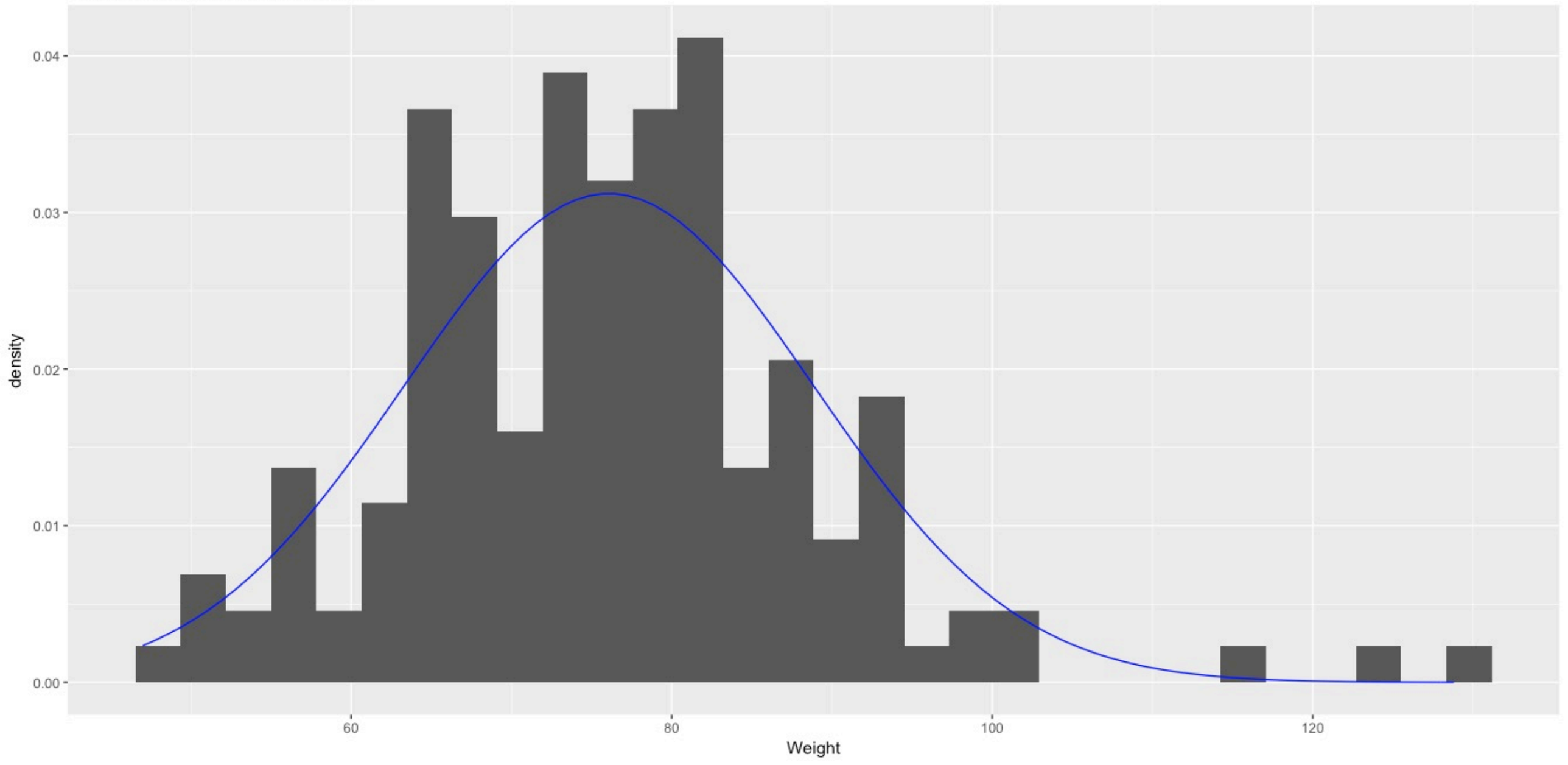

Normal distribution of BMI in all the students
Certification Program on Business Analytics
density
0.15
0.10
0.05
0.00
20
25
30
35
40
Height

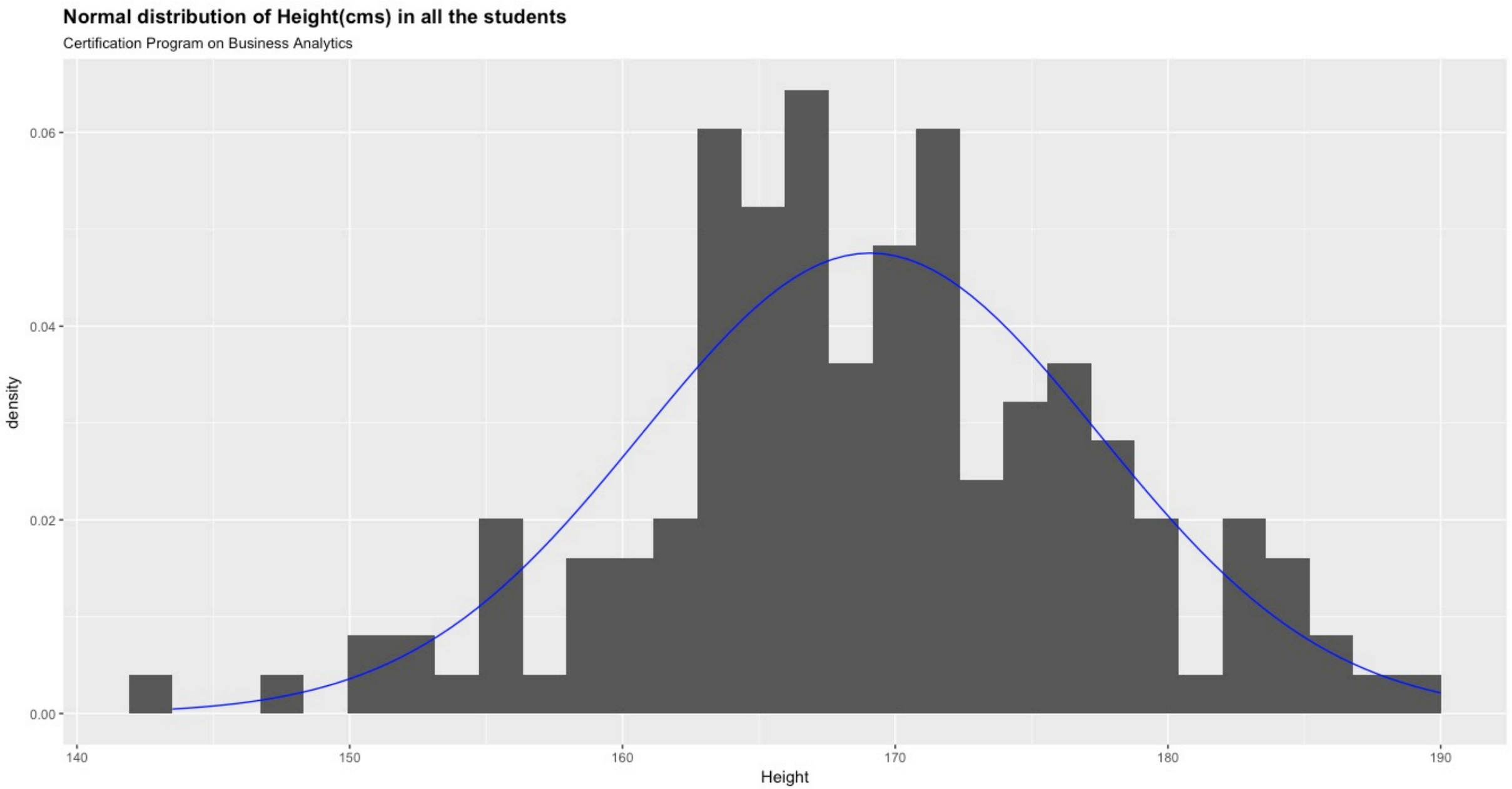

Normal distribution of Height(cms) in all the students
Certification Program on Business Analytics
0.06
0.04
0.02
0.00
density
140
150
160
170
180
190
Height


Normal distribution of Height(cms) in all the students
Certification Program on Business Analytics
0.06
0.04
0.02
0.00
density
140
150
160
170
180
190
Height

Normal distribution of Weight(kgs) in all the students

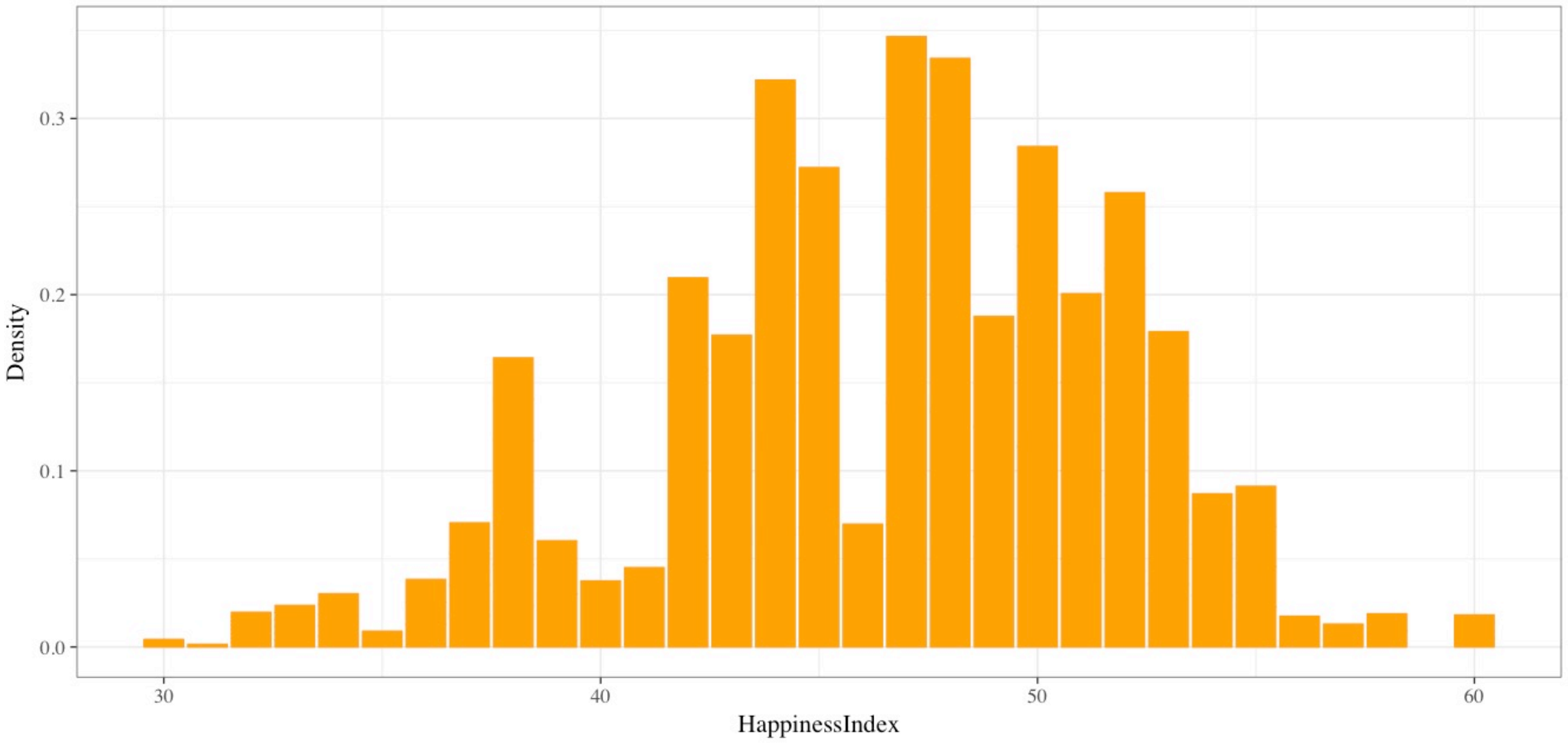
Certification Program on Business Analytics

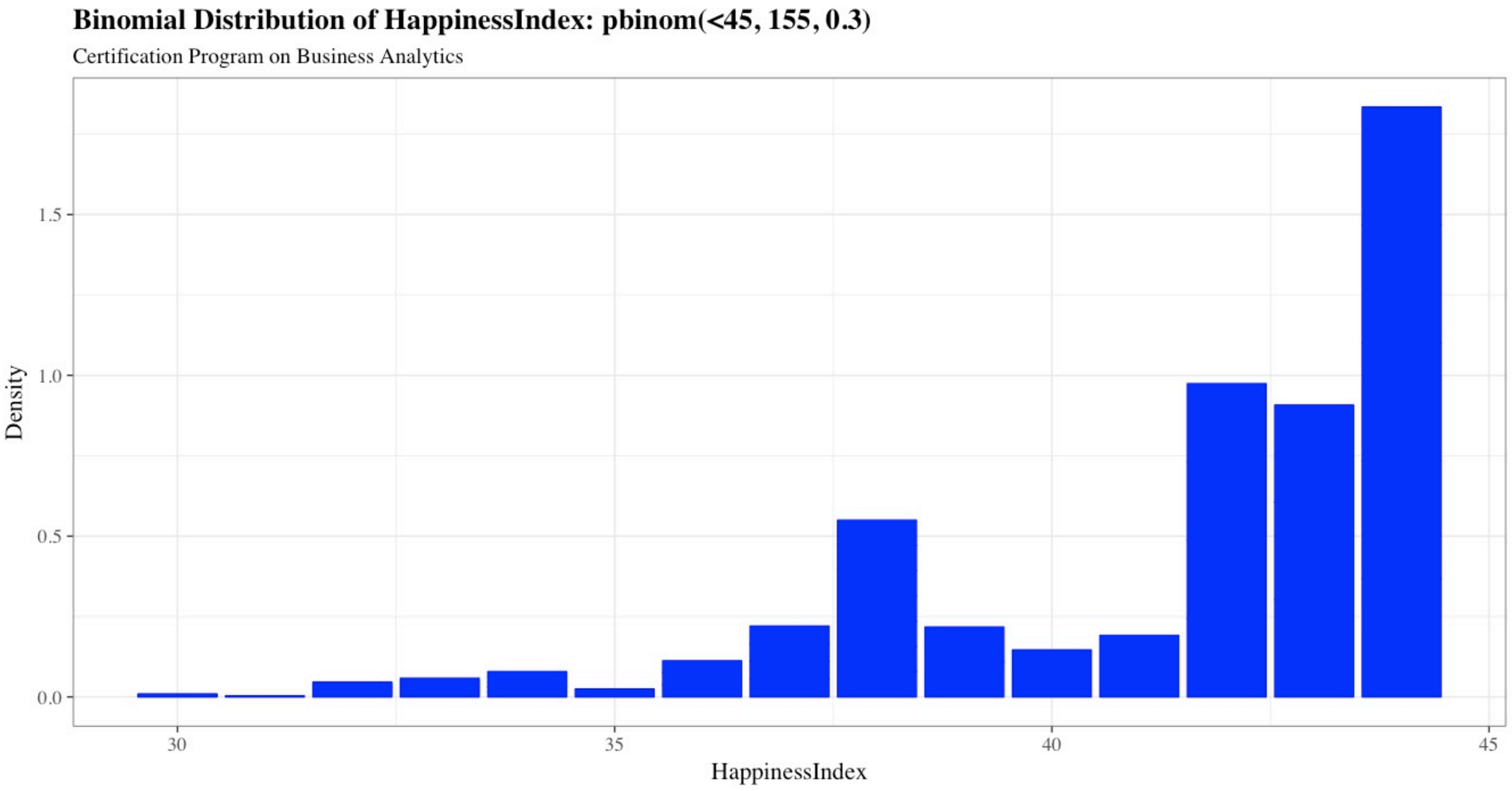



Normal distribution of Height(cms) in all the students
Certification Program on Business Analytics
0.06
0.04
0.02
0.00
density
140
150
160
170
180
190
Height

Binomial Distribution of HappinessIndex: dbinom(30:60, 155, 0.3)

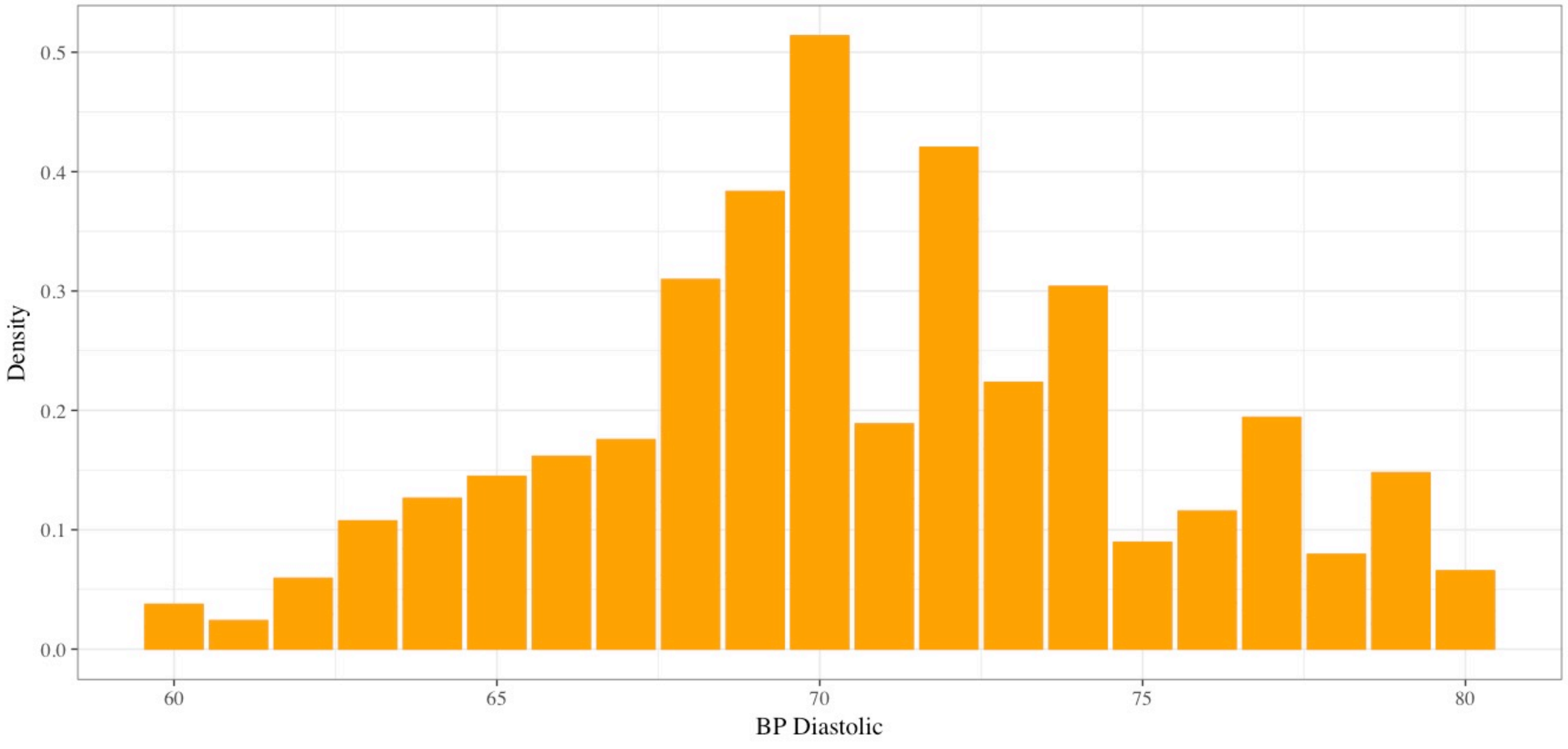
Certification Program on Business Analytics




Binomial Distribution of HappinessIndex: pbinom(<45, 155, 0.3)
Certification Program on Business Analytics
Density
1.5
1.0
0.5
0.0
30
35
40
45
HappinessIndex

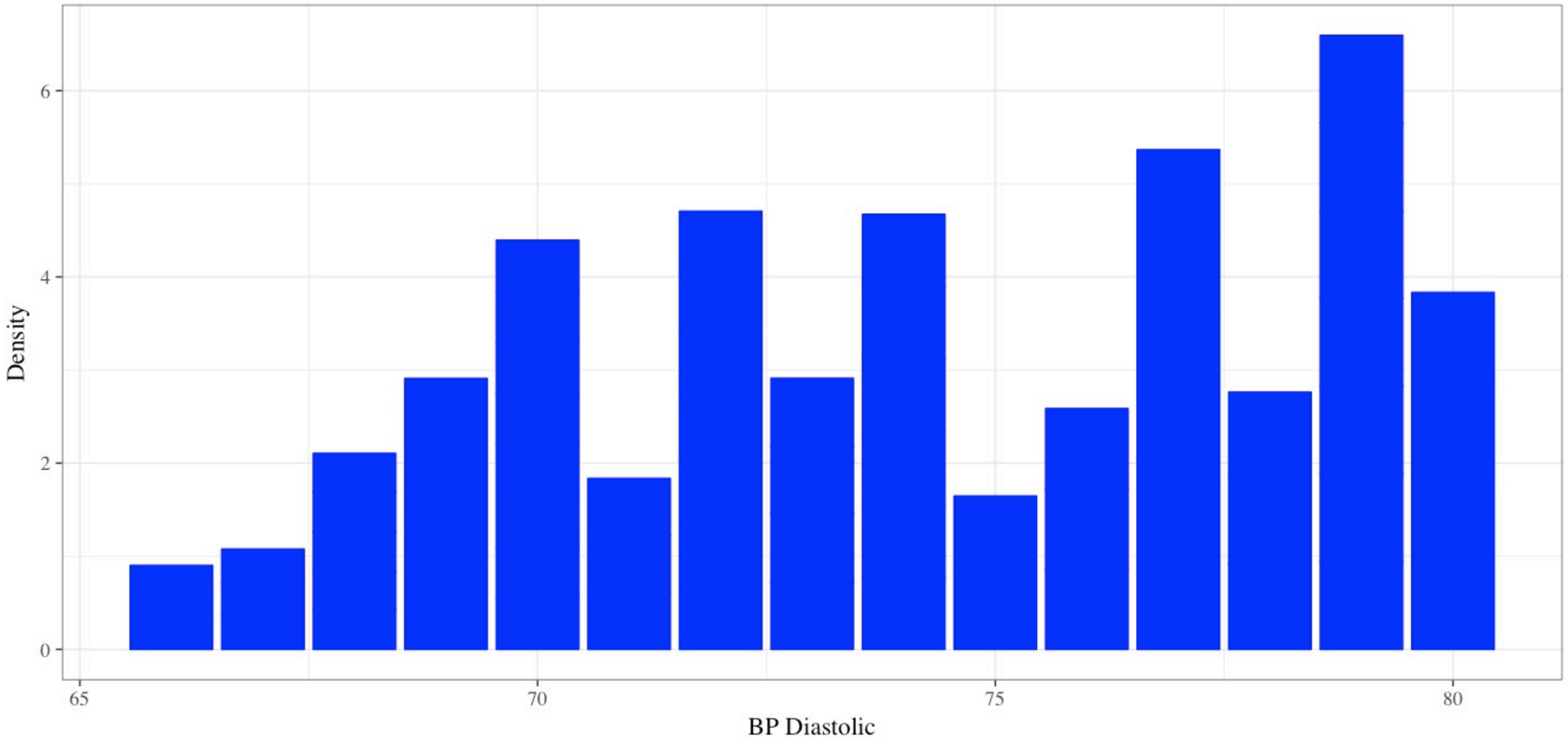
Binomial Distribution of BP Diastolic: dbinom(60:80, 155, 0.45)

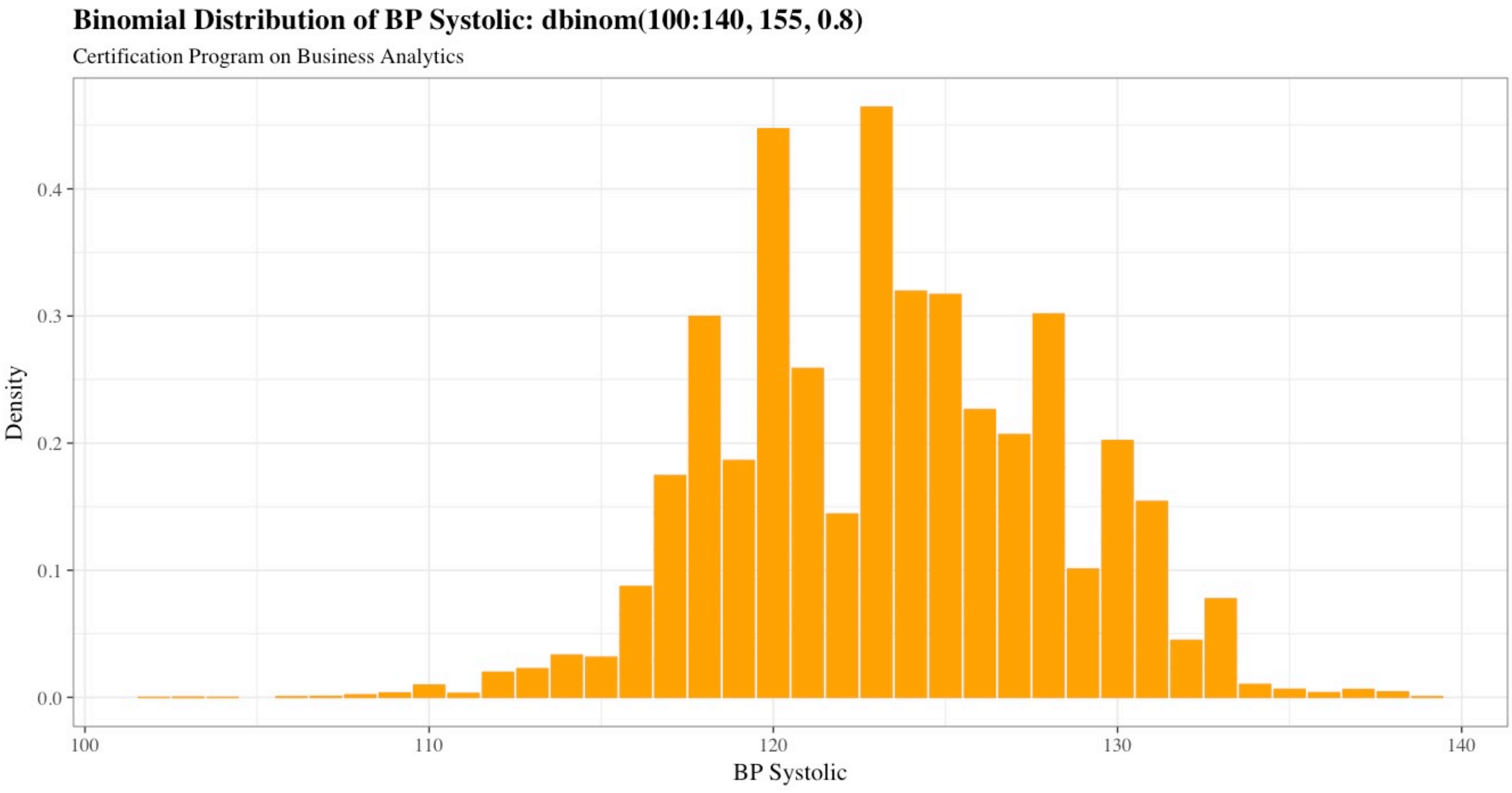
Certification Program on Business Analytics



Binomial Distribution of BP Diastolic: pbinom(>65, 155, 0.45)

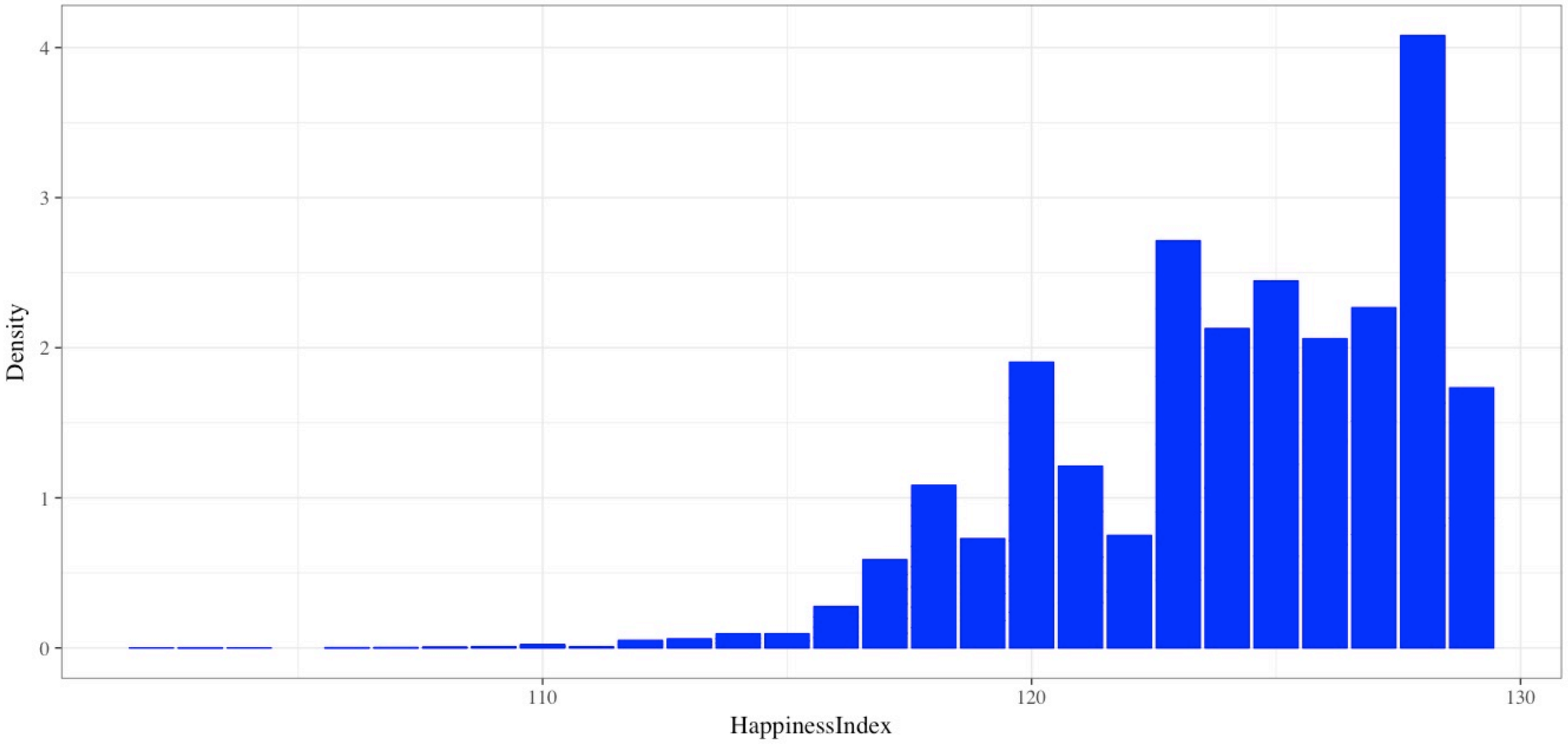
Certification Program on Business Analytics

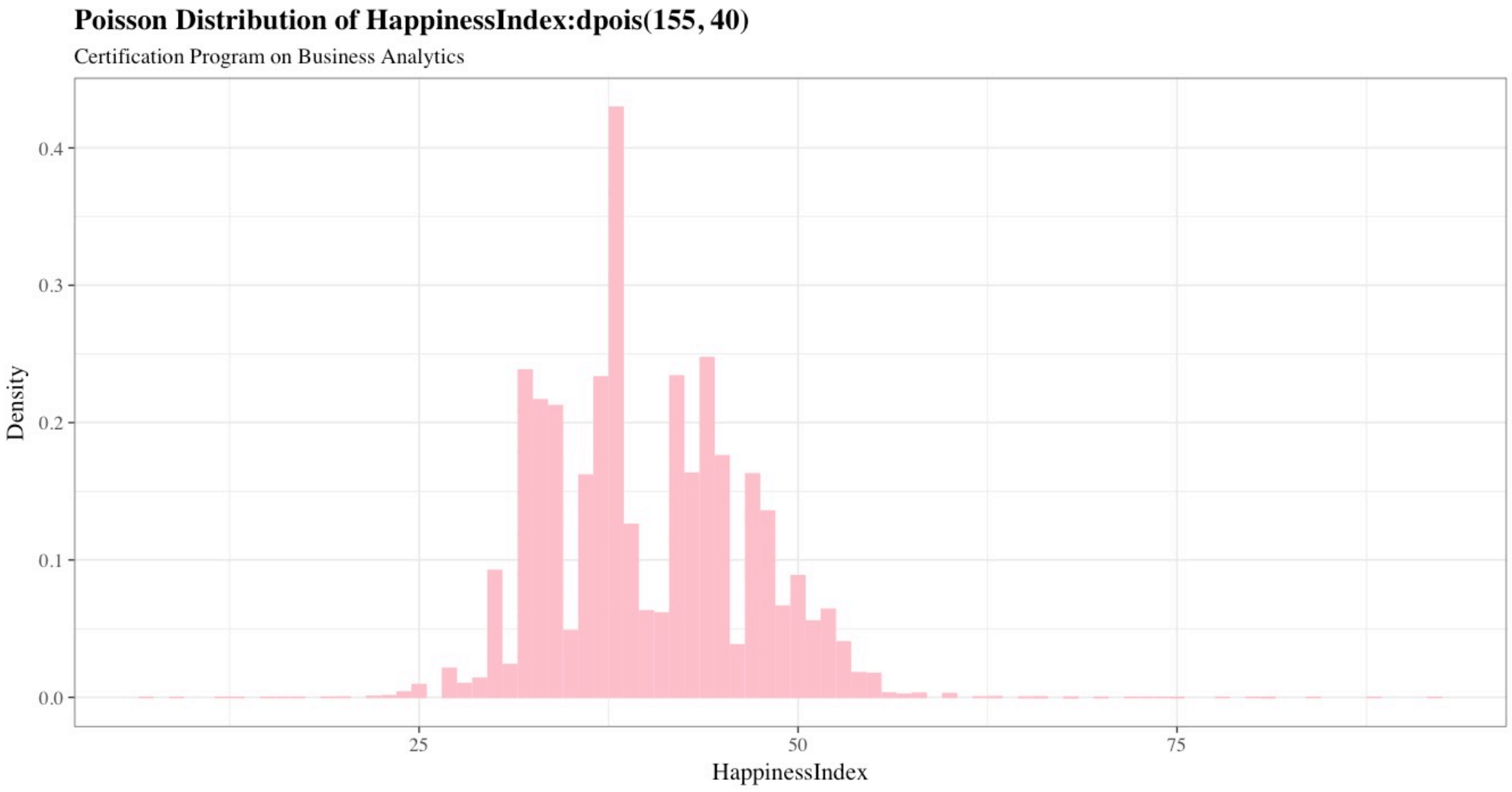


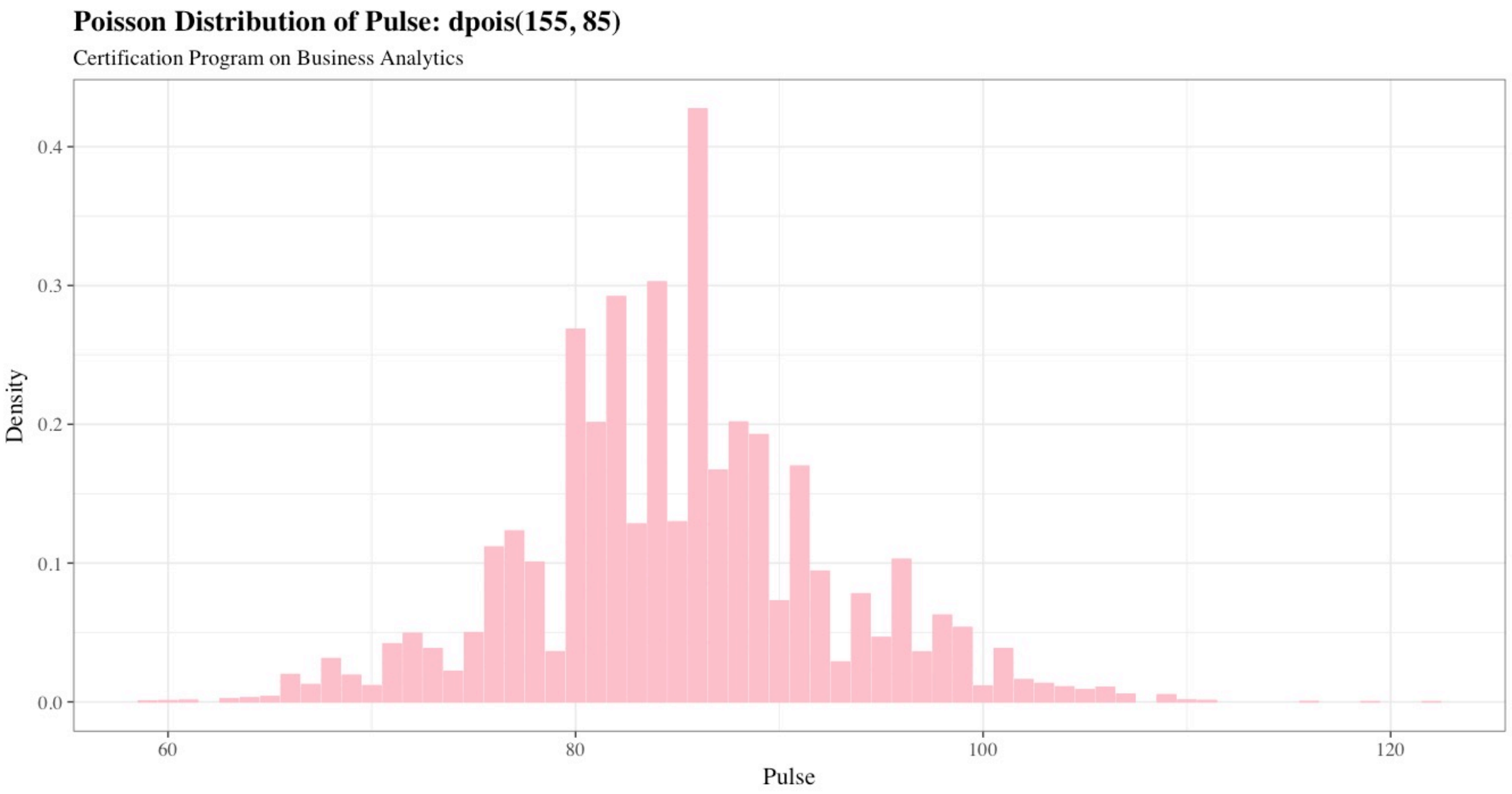

Binomial Distribution of BP Systolic: dbinom(100:140, 155, 0.8)
Certification Program on Business Analytics
Density
0.0
0.1
0.2
0.3
0.4
100
110
120
130
140
BP Systolic

Binomial Distribution of Systolic: pbinom(<130, 155, 0.8)

Certification Program on Business Analytics

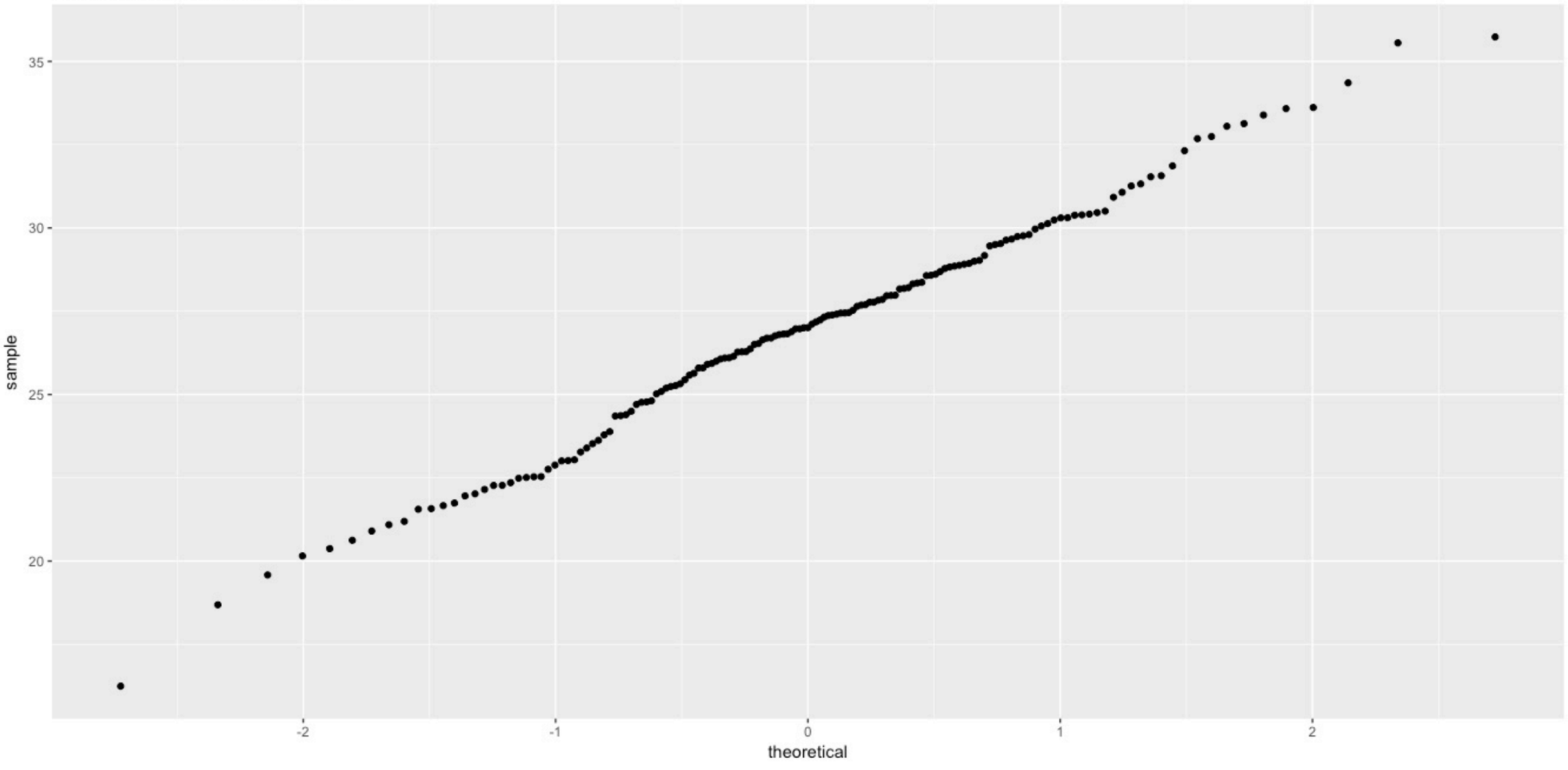


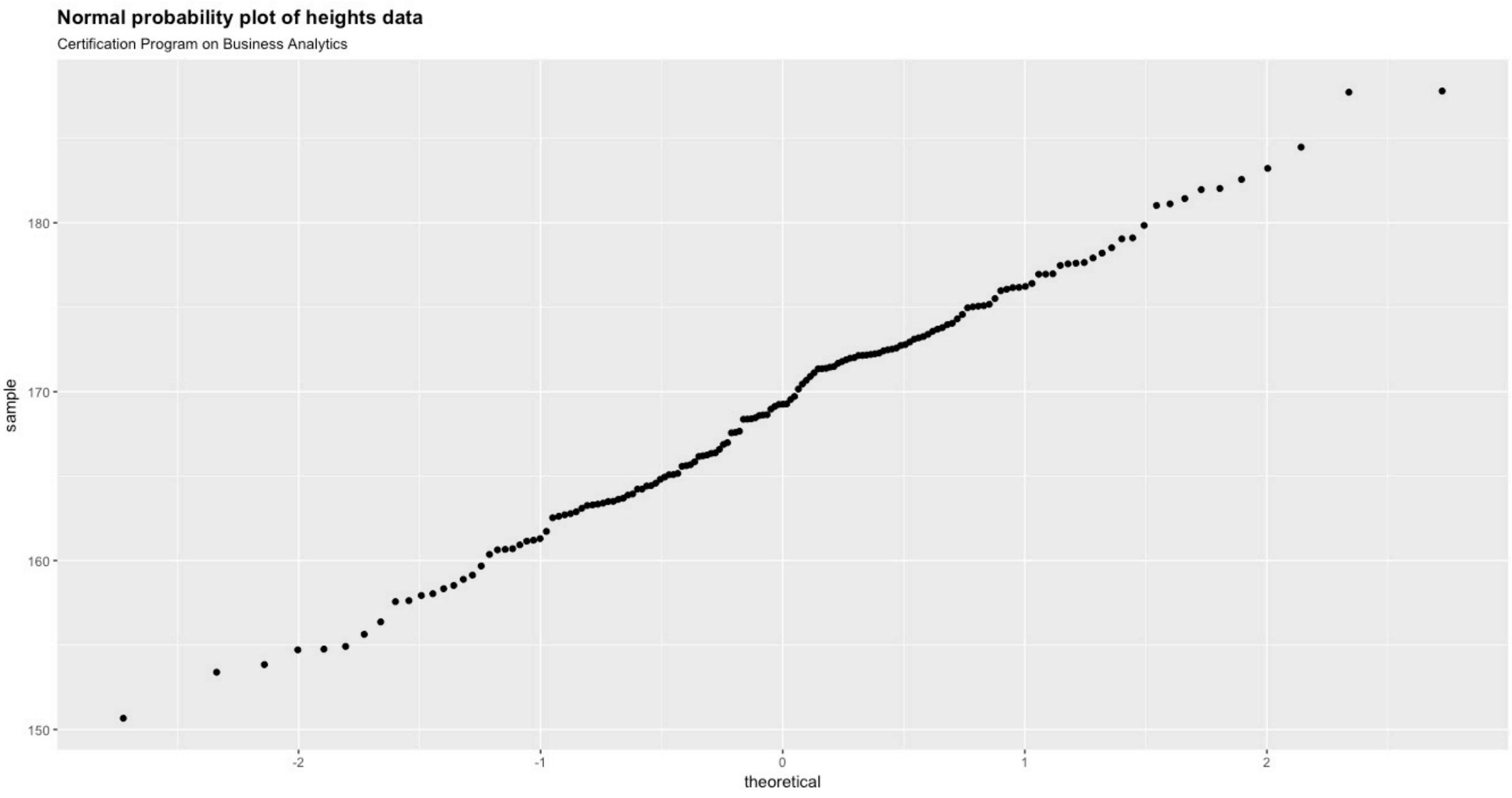

Poisson Distribution of HappinessIndex:dpois(155, 40)
Certification Program on Business Analytics
Density
0.4
0.3
0.2
0.1
0.0
25
50
75
HappinessIndex


Poisson Distribution of Pulse: dpois(155, 85)
Certification Program on Business Analytics
Density
0.4
0.3
0.2
0.1
0.0
60
80
100
120
Pulse

Normal probability plot of BMI data

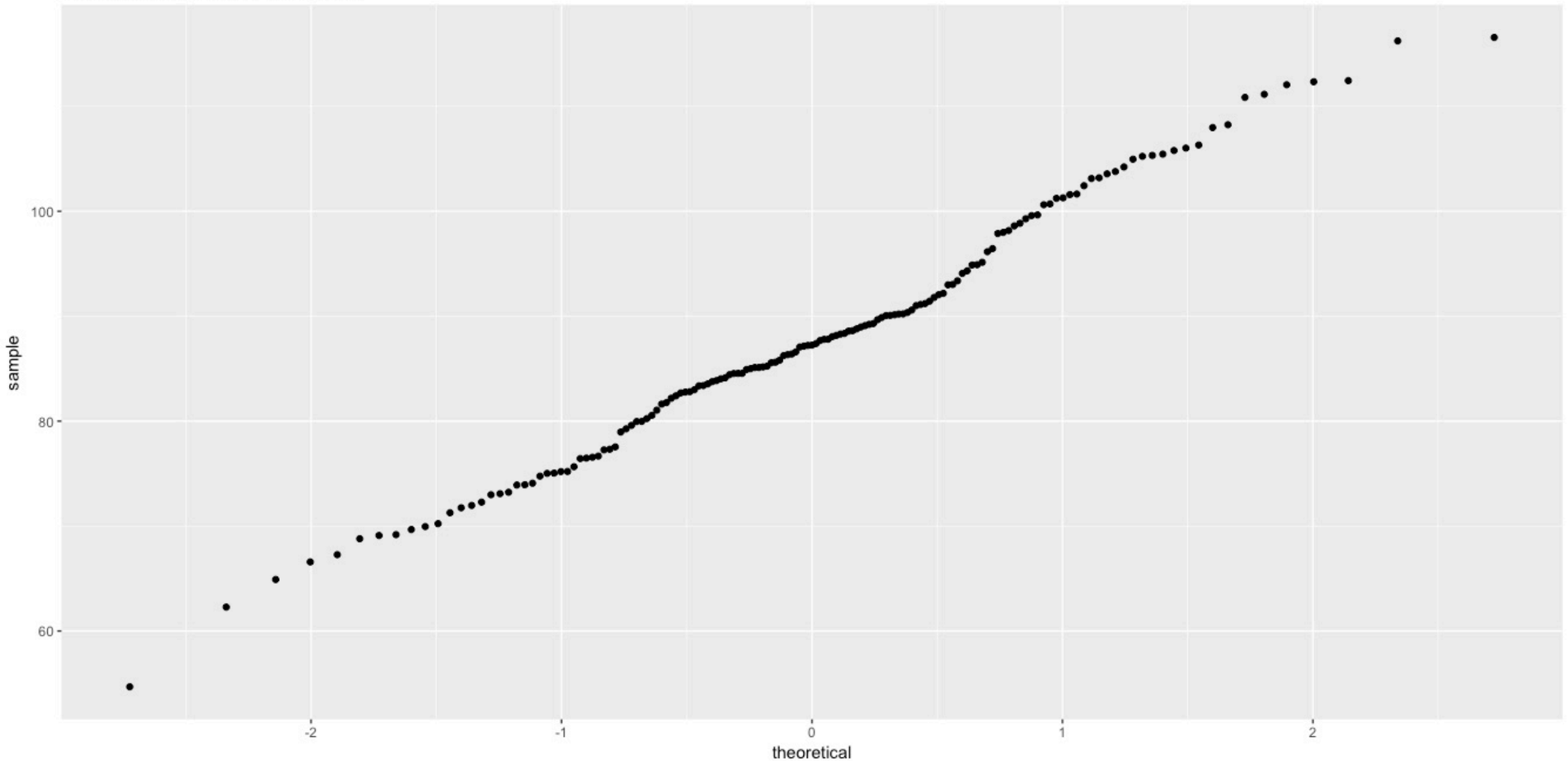
Certification Program on Business Analytics

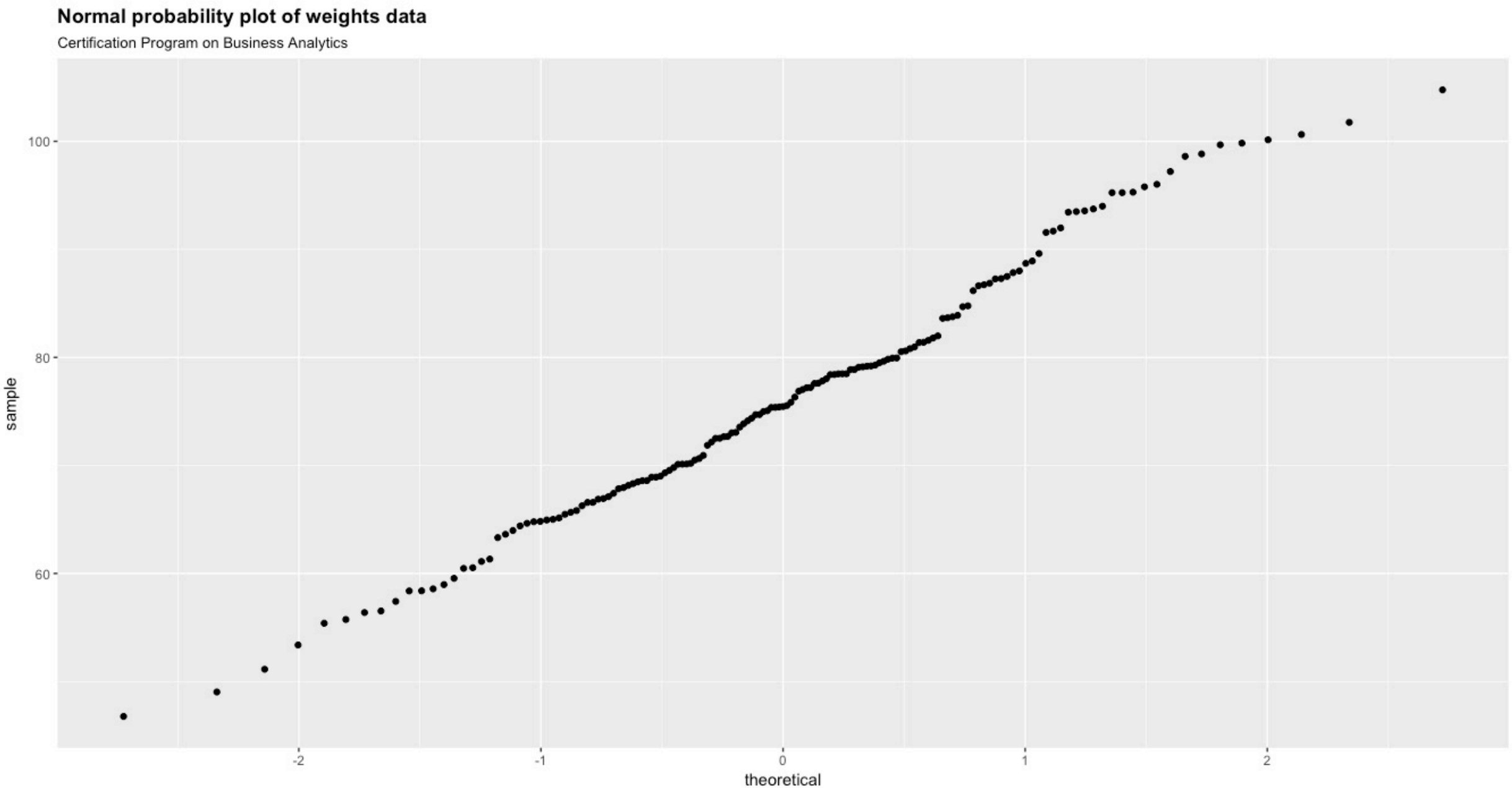



Normal probability plot of heights data
Certification Program on Business Analytics
sample
180
170
160
150
-2
-1
0
1
2
theoretical

Normal probability plot of simulated pulse rates data

Certification Program on Business Analytics




Normal probability plot of weights data
Certification Program on Business Analytics
100
80
60
sample
-2
-1
0
1
2
theoretical

CBS book problems have been solved manually in a paper and documented in PDF, it's attached to this separately.

Mallesham Yamulla