A Primer for Statistical Tests

R for Beginners

-

basic statistics for biologists

Erik Kusch
PhD Student
Aarhus University
Department of Bioscience
Section for Ecoinformatics & Biodiversity
Center for Biodiversity and Dynamics in a Changing World (BIOCHANGE)
Ny Munkegade 116, Building 1540
8000 Aarhus
Denmark
email: erik@i-solution.de

Summary:

These are the solutions to the exercises contained within the handout to "A Primer For Statistical Tests" which walks you through the basics of variables, their scales and distributions. Keep in mind that there is probably a myriad of other ways to reach the same conclusions as presented in these solutions.

Contents

1. Loading the R Environment Object

```
load("Data/Primer.RData") # load data file from Data folder
```

2. Variables

2.1 Finding Variables

```
ls() # list all elements in working environment
```

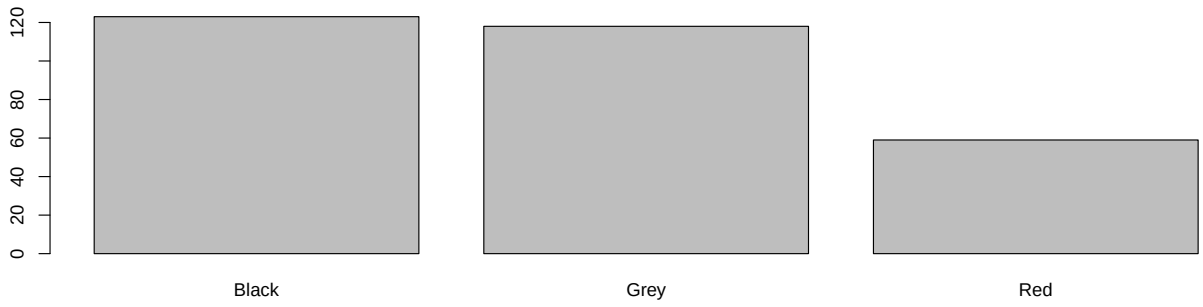
```
## [1] "Colour"              "Depth"               "IndividualsPassingBy"
## [4] "Length"              "Reproducing"         "Sex"
## [7] "Size"                "Temperature"
```

2.2 Colour

```
class(Colour) # mode
```

```
## [1] "character"
```

```
barplot(table(Colour)) # fitting?
```



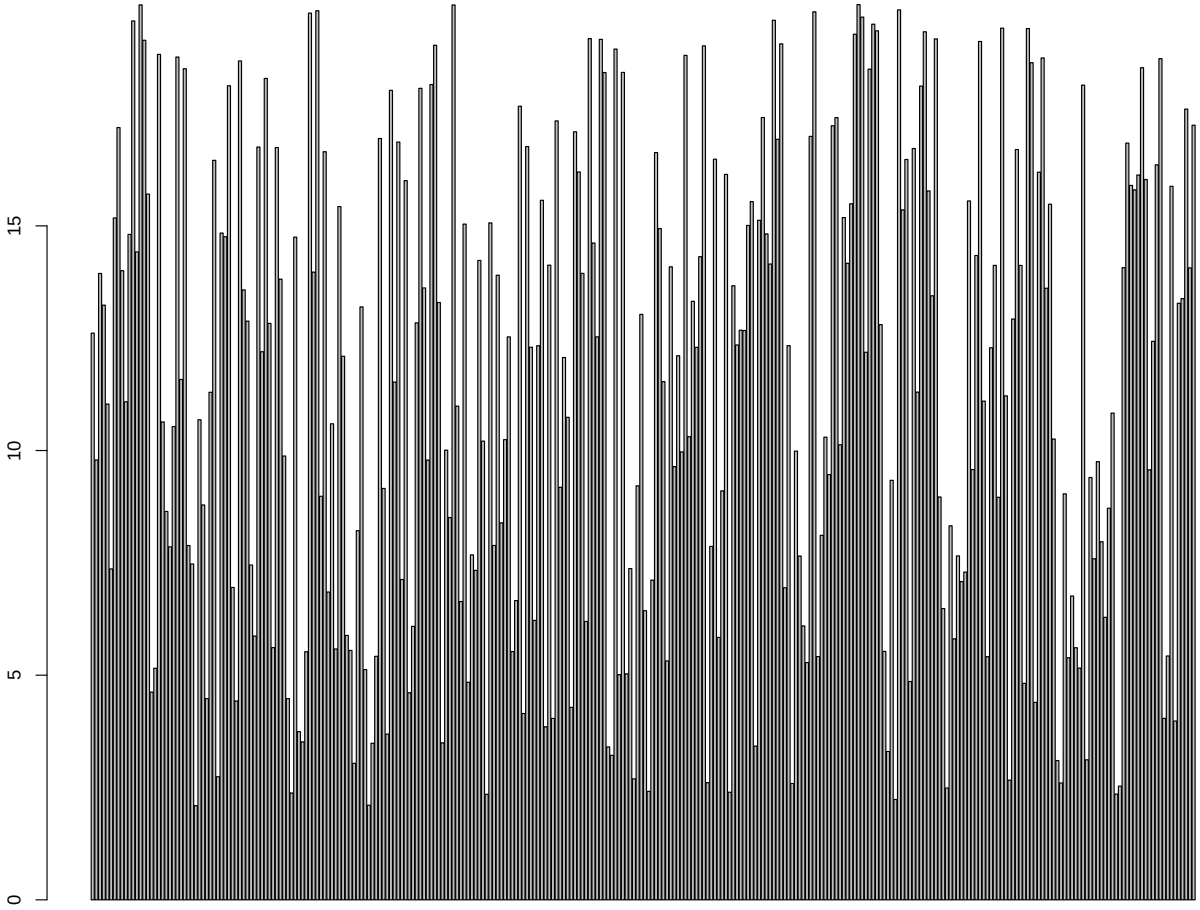
Question	Answer
Mode?	character
Which scale?	Nominal
What's implied?	Categorical data that can't be ordered
Does data fit scale?	Yes

2.3 Depth

```
class(Depth) # mode
```

```
## [1] "numeric"
```

```
barplot(Depth) # fitting?
```



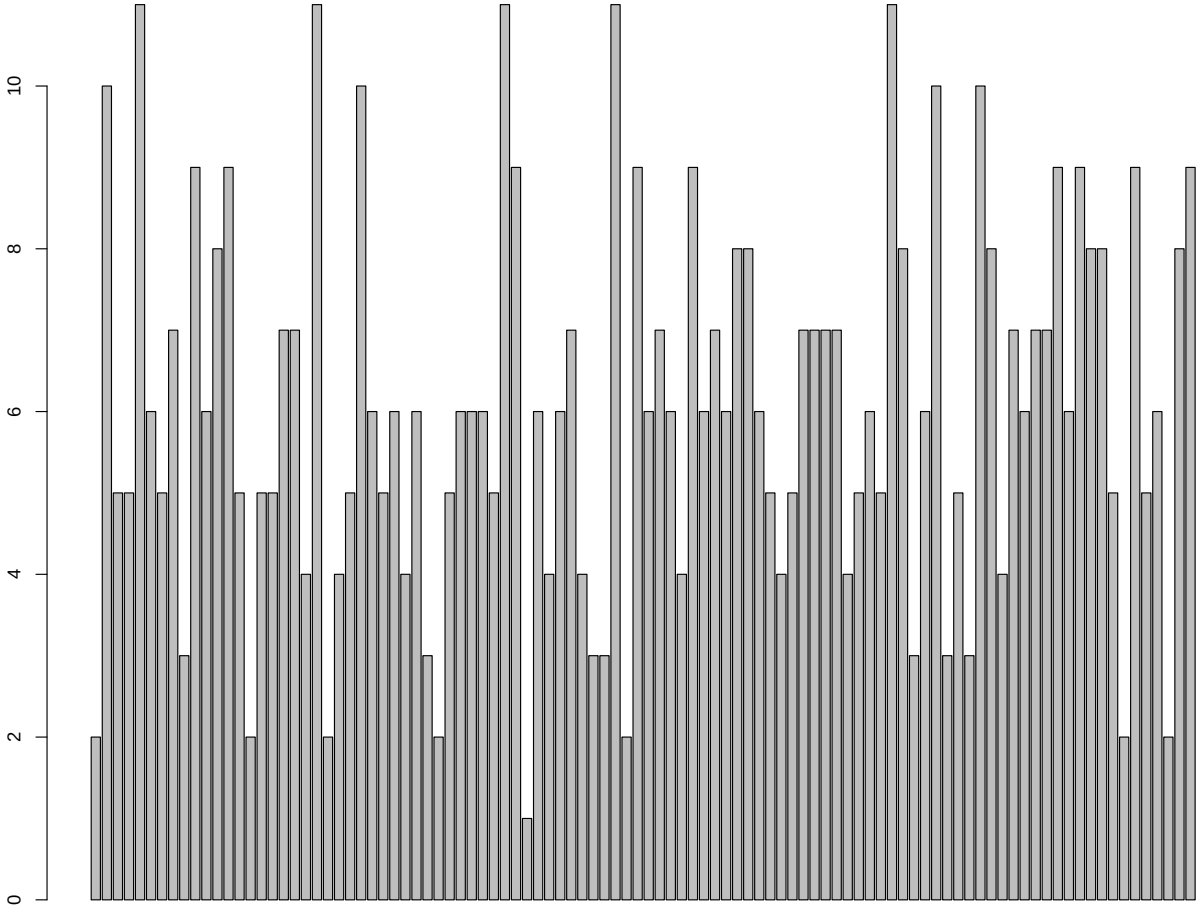
Question	Answer
Mode?	numeric
Which scale?	Interval/Discrete
What's implied?	Continuous data with a non-absence point of origin
Does data fit scale?	Debatable (is 0 depth absence of depth?)

2.4 IndividualsPassingBy

```
class(IndividualsPassingBy) # mode
```

```
## [1] "integer"
```

```
barplot(IndividualsPassingBy) # fitting?
```



Question	Answer
Mode?	integer
Which scale?	Integer
What's implied?	Only integer numbers with an absence point of origin
Does data fit scale?	Yes

2.5 Length

```
class(Length) # mode
```

```
## [1] "numeric"
```

```
barplot(Length) # fitting?
```

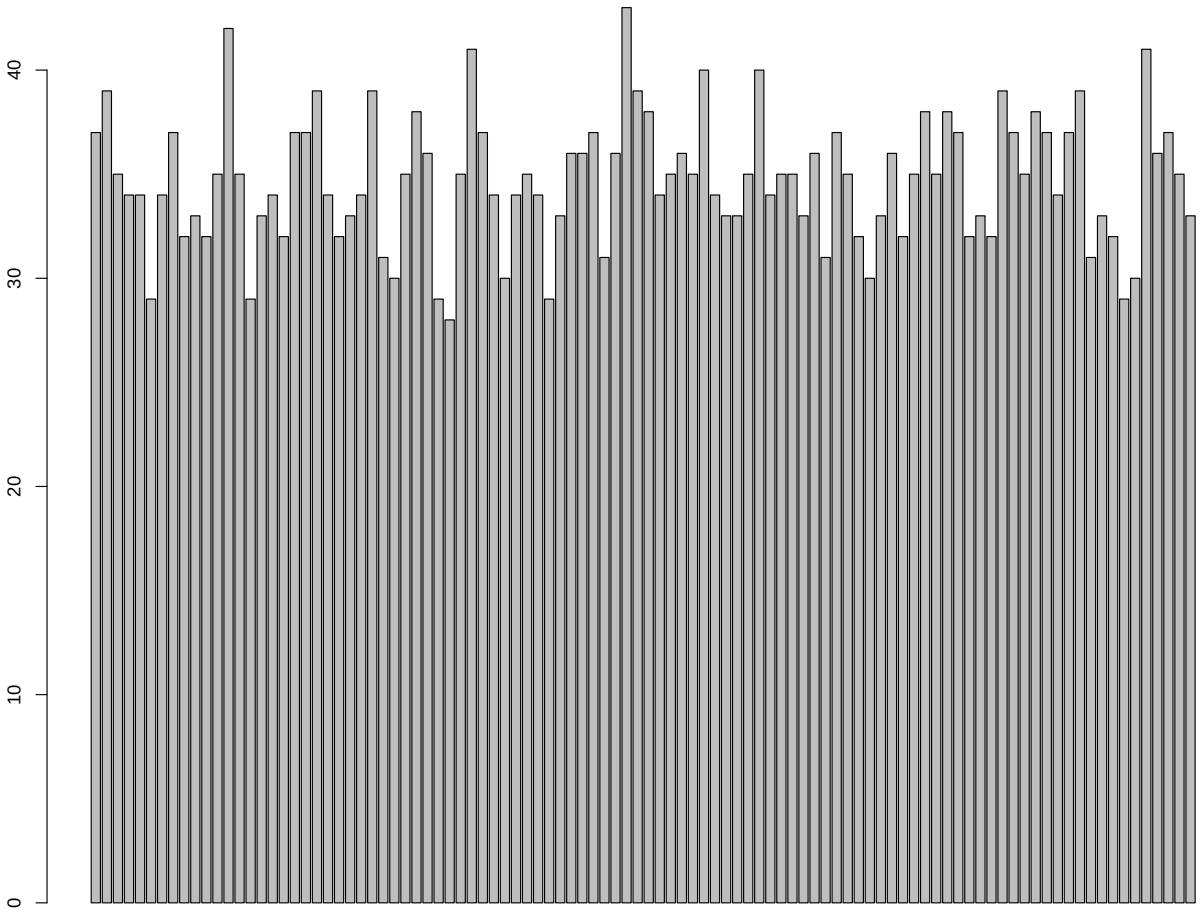
Question	Answer
Mode?	numeric
Which scale?	Relation/Ratio
What's implied?	Continuous data with an absence point of origin
Does data fit scale?	Yes

2.6 Reproducing

```
class(Reproducing) # mode
```

```
## [1] "integer"
```

```
barplot(Reproducing) # fitting?
```



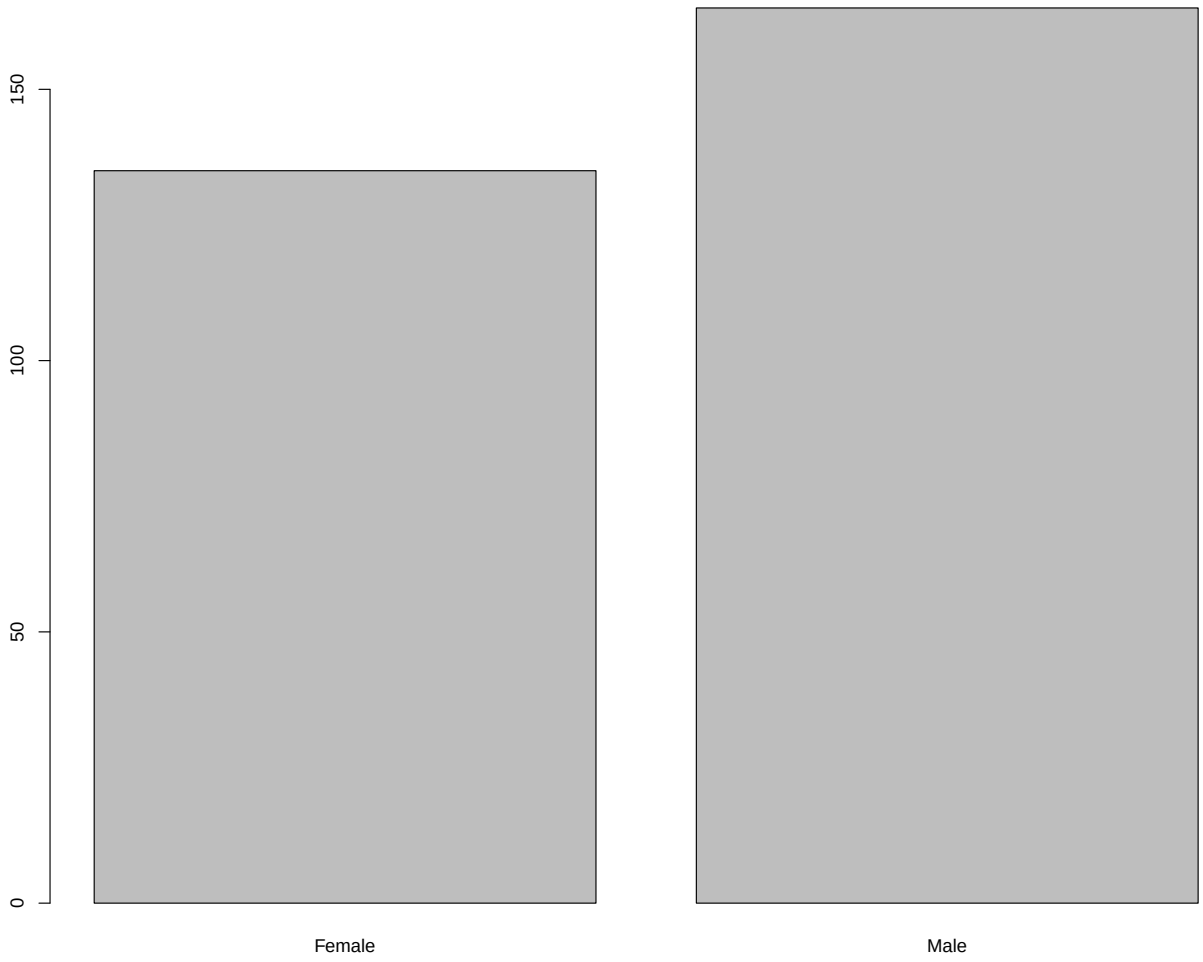
Question	Answer
Mode?	integer
Which scale?	Integer
What's implied?	Only integer numbers with an absence point of origin
Does data fit scale?	Yes

2.7 Sex

```
class(Sex) # mode
```

```
## [1] "factor"
```

```
barplot(table(Sex)) # fitting?
```



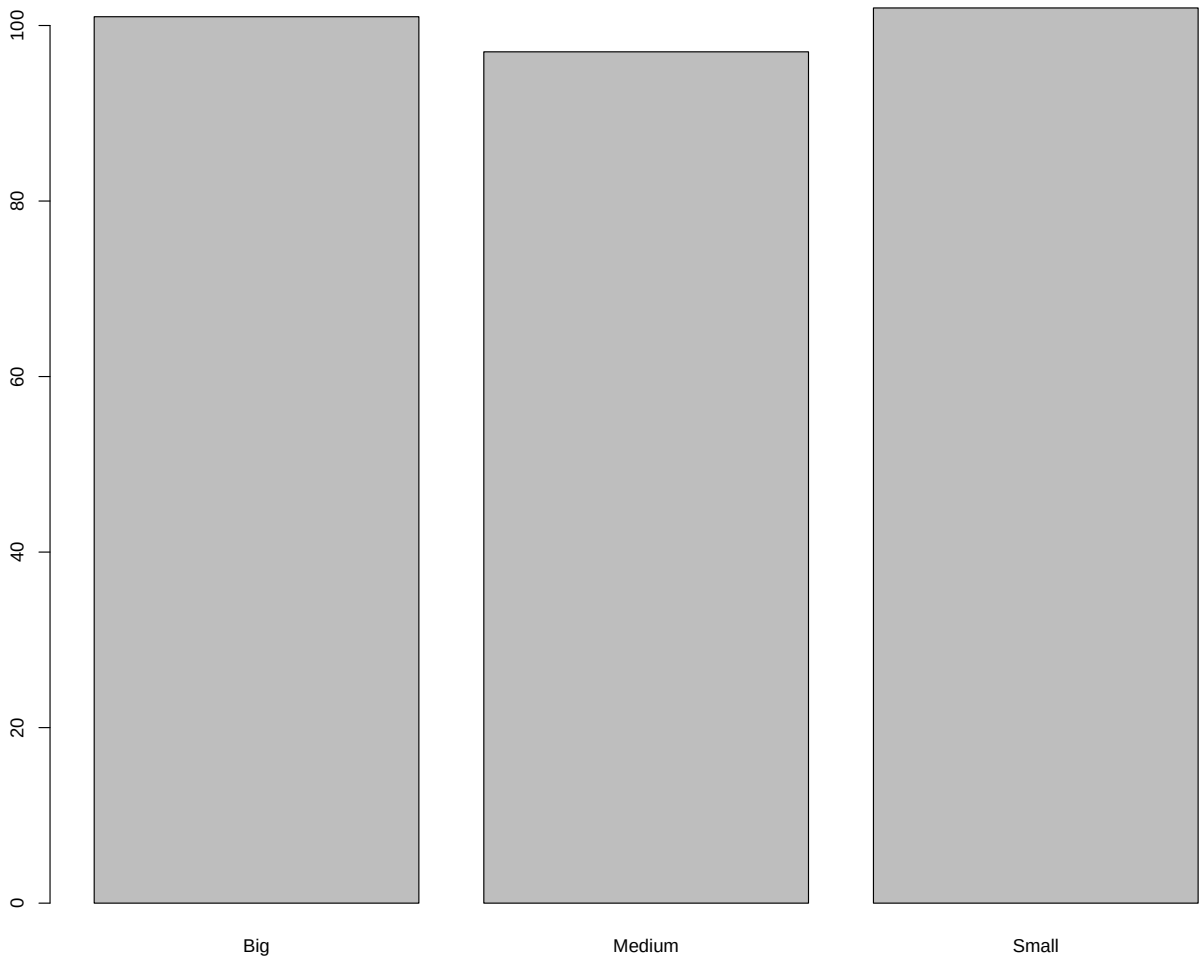
Question	Answer
Mode?	factor
Which scale?	Binary
What's implied?	Only two possible outcomes
Does data fit scale?	Yes

2.8 Size

```
class(Size) # mode
```

```
## [1] "character"
```

```
barplot(table(Size)) # fitting?
```



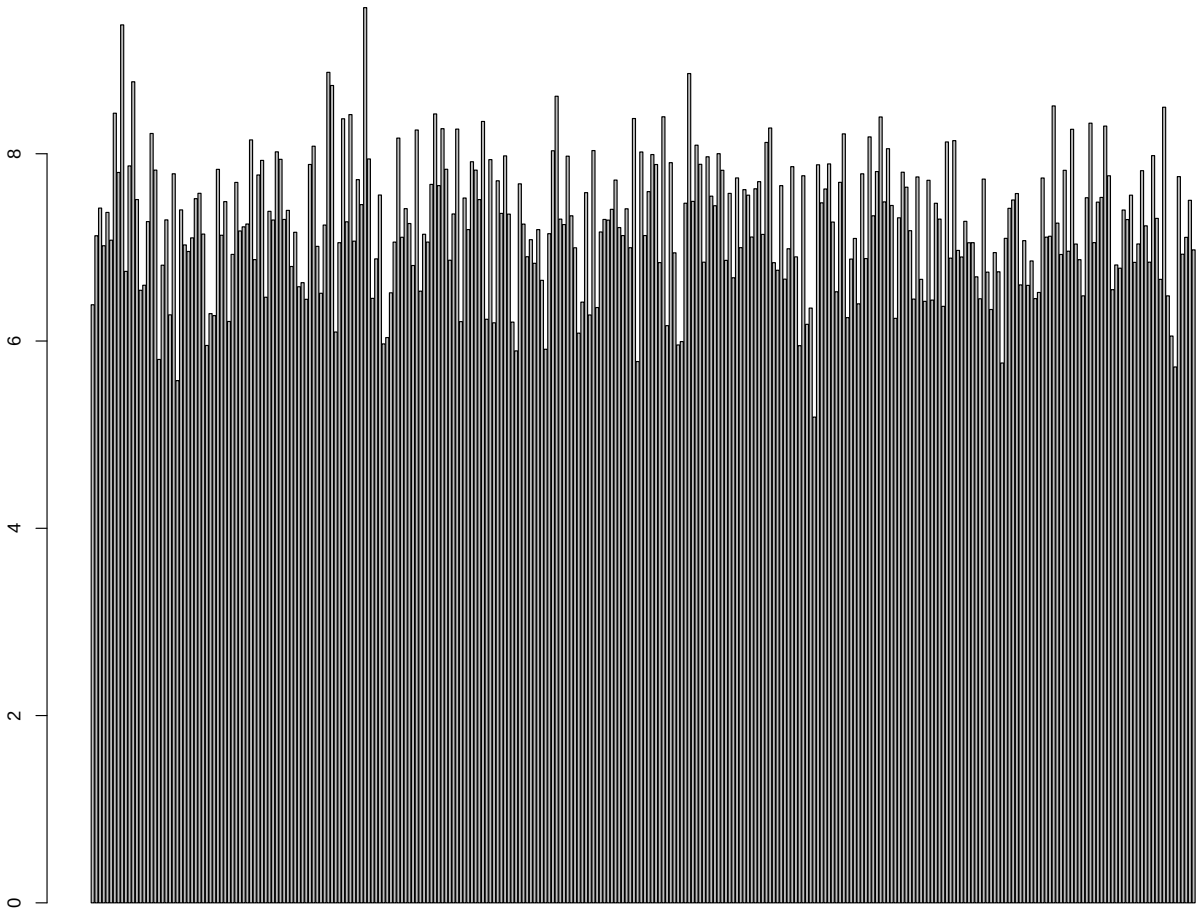
Question	Answer
Mode?	character
Which scale?	Ordinal
What's implied?	Categorical data that can be ordered
Does data fit scale?	Yes

2.9 Temperature

```
class(Temperature) # mode
```

```
## [1] "numeric"
```

```
barplot(Temperature) # fitting?
```

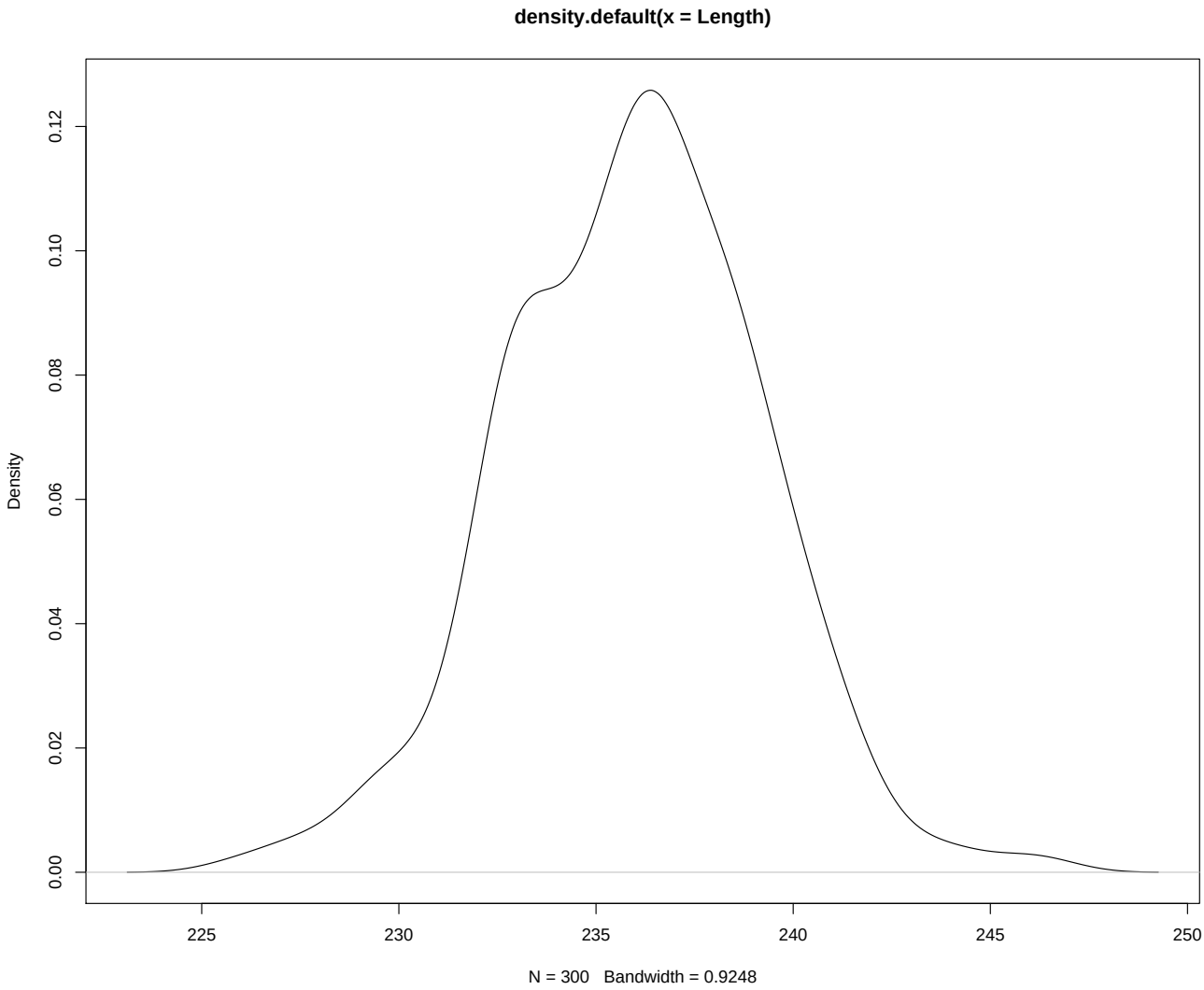


Question	Answer
Mode?	numeric
Which scale?	Interval/Discrete
What's implied?	Continuous data with a non-absence point of origin
Does data fit scale?	Yes (the data is clearly recorded in degree Celsius)

3. Distributions

3.1 Length

```
plot(density(Length)) # distribution plot
```



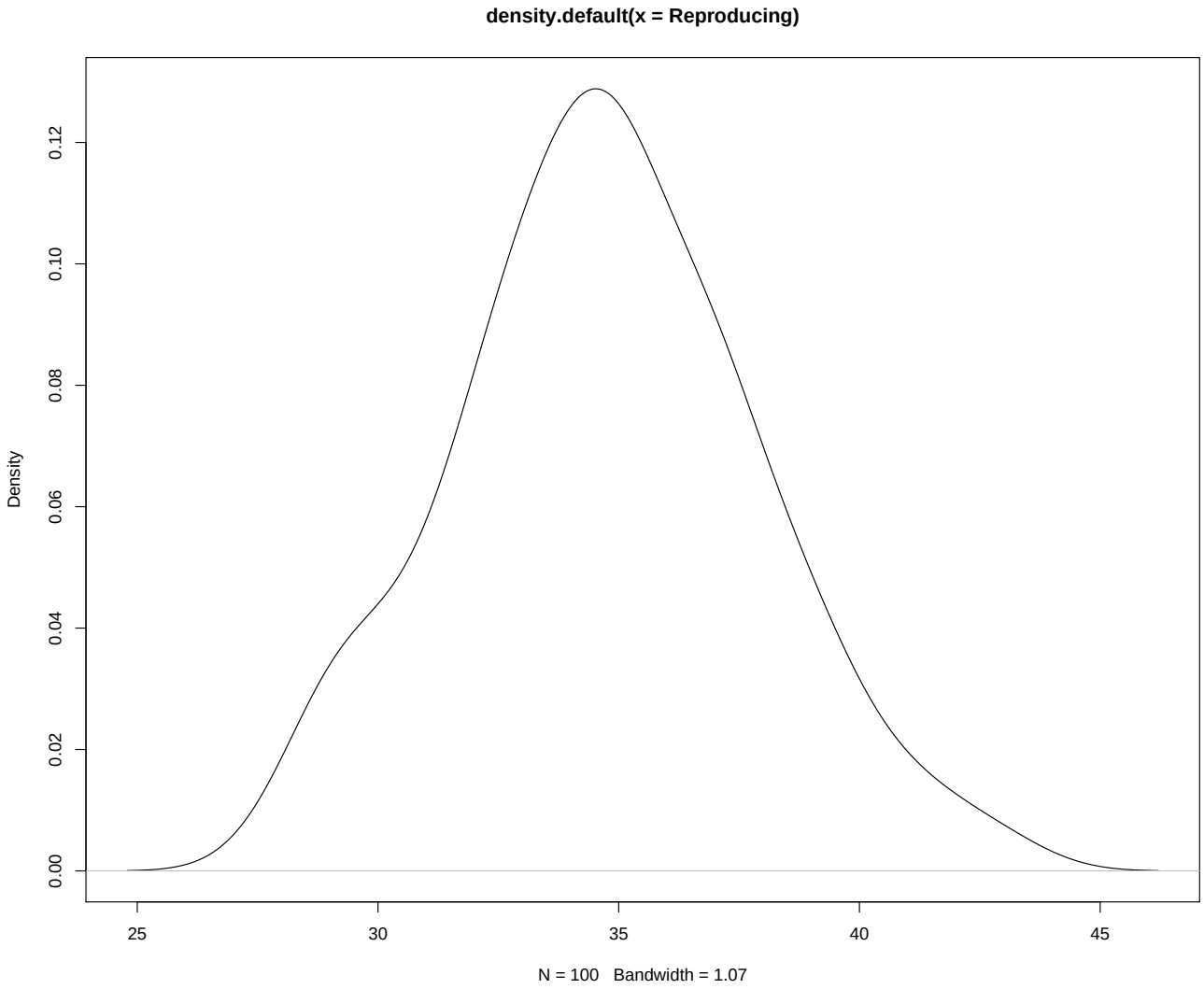

```
shapiro.test(Length) # normality check
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Length
## W = 1, p-value = 0.4
```

The data is **normal distributed**.

3.2 Reproducing

```
plot(density(Reproducing)) # distribution
```



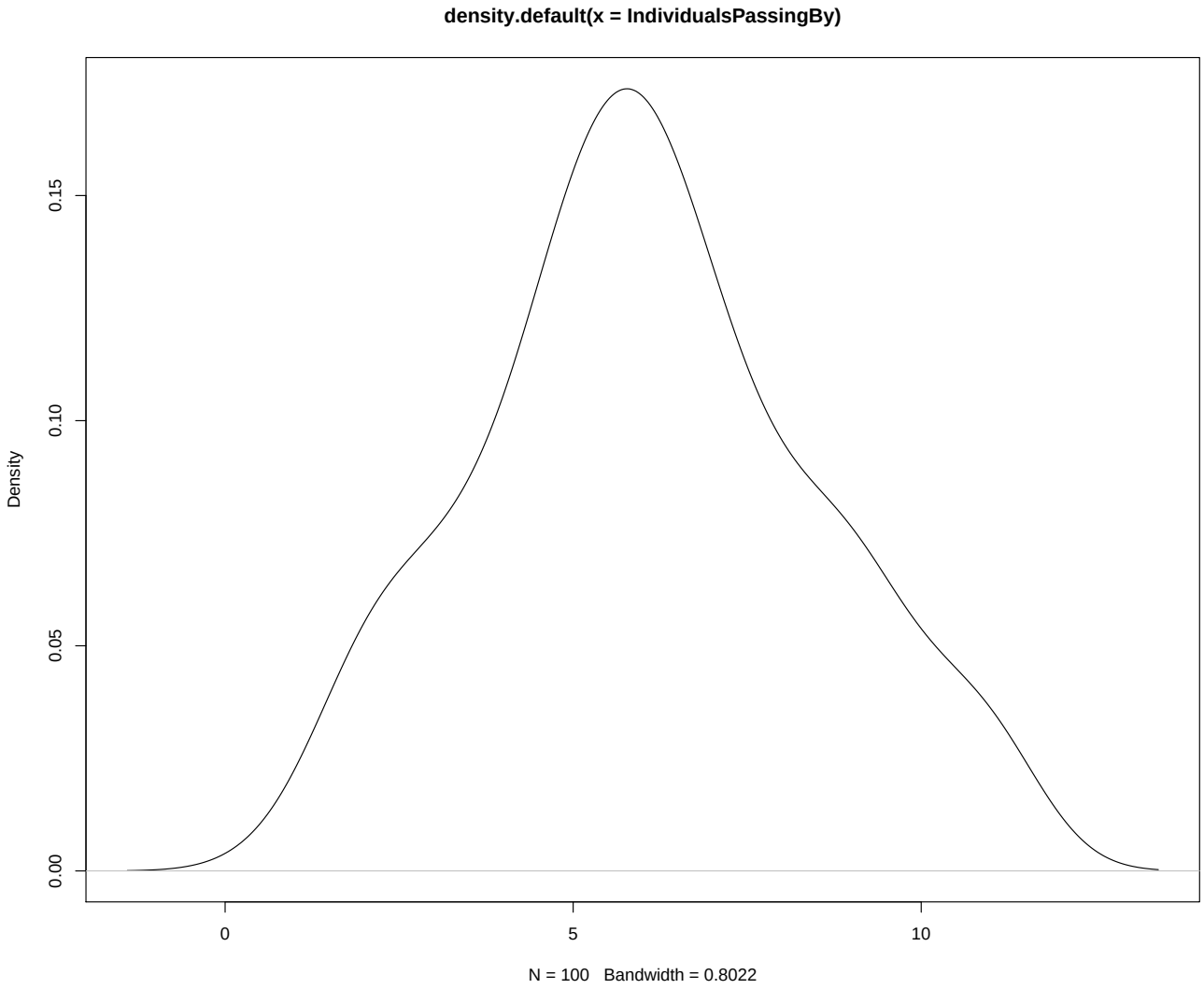

```
shapiro.test(Reproducing) # normality check
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Reproducing
## W = 1, p-value = 0.3
```

The data is **binomial distributed** (i.e. "How many individuals manage to reproduce") but looks **normal distributed**. The normal distribution doesn't make sense here because it implies continuity whilst the data only comes in integers.

3.3 IndividualsPassingBy

```
plot(density(IndividualsPassingBy)) # distribution
```



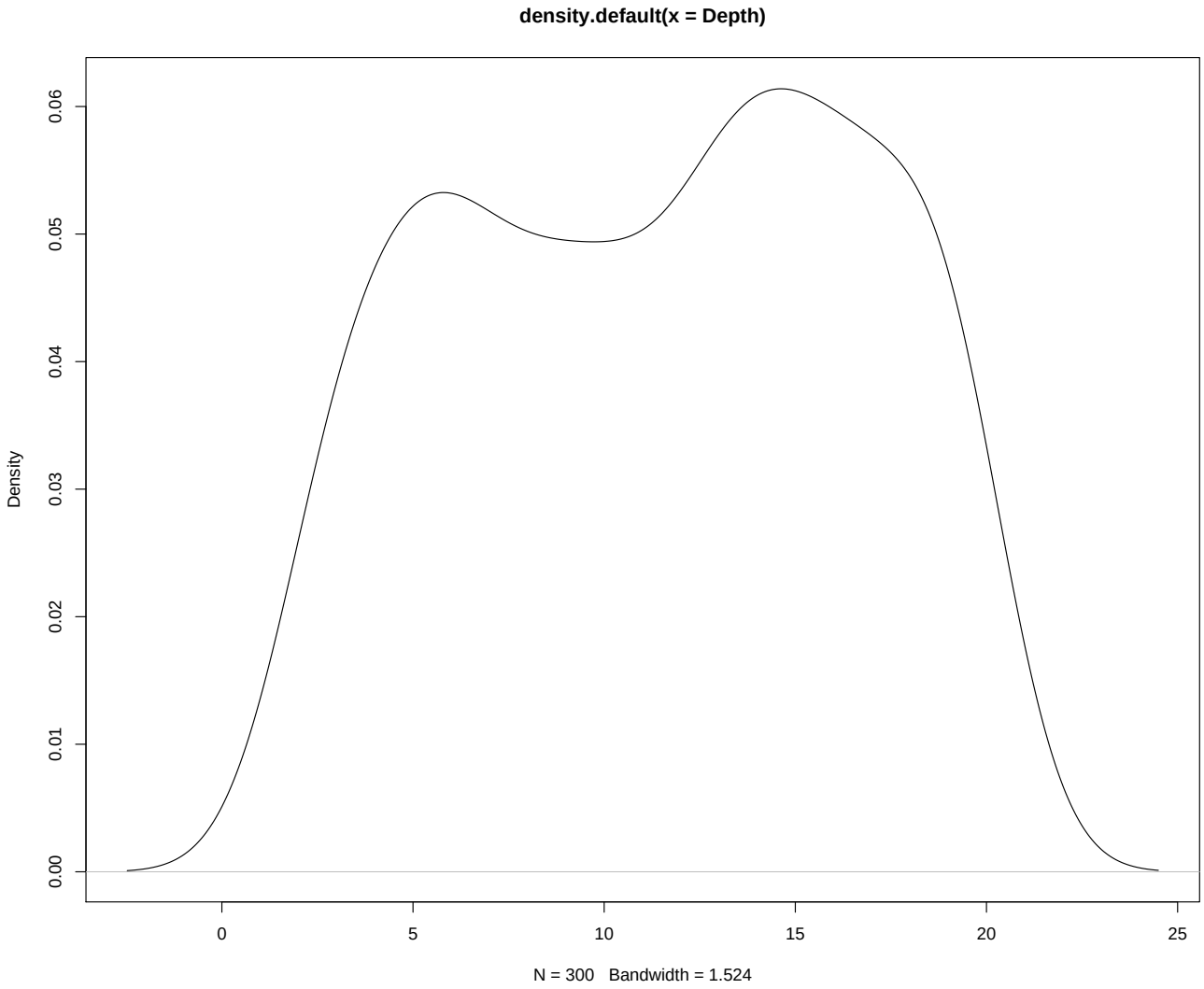

```
shapiro.test(IndividualsPassingBy) # normality check
```

```
##
##  Shapiro-Wilk normality test
##
## data:  IndividualsPassingBy
## W = 1, p-value = 0.02
```

The data is **poisson distributed** (i.e. “How many individuals pass by an observer in a given time frame?”).

3.4 Depth

```
plot(density(Depth)) # distribution
```



The data is **uniform distributed**. You don't know this distribution class from the lectures and I only wanted to confuse you with this to show you that there's much more out there than I can show in our lectures.