



[Course](#) > [Section...](#) > [1.2 Intr...](#) > [Descri...](#)

Audit Access Expires Mar 24, 2020

You lose all access to this course, including your progress, on Mar 24, 2020.

Upgrade by Feb 18, 2020 to get unlimited access to the course as long as it exists on the site. [Upgrade now](#)

Describe Heights to ET

Describe Heights to ET

[Start of transcript. Skip to the end.](#)



RAFAEL IRIZARRY: Here we introduce a new motivating problem.

It is an artificial one, but it'll help us illustrate the concepts needed

to understand distributions.

Later, we'll move on to more realistic examples.

Pretend that we have to

Video



Download video file

Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

Textbook links

This video corresponds to the following sections:

- [textbook case study of describing student heights](#)
- [textbook section on the distribution function](#)
- [textbook section on the cumulative distribution function](#)
- [textbook section on histograms](#)

Key points

- A distribution is a function or description that shows the possible values of a variable and how often those values occur.
- For categorical variables, the distribution describes the proportions of each category.
- A *frequency table* is the simplest way to show a categorical distribution. Use `prop.table()` to convert a table of counts to a frequency table. *Barplots* display the distribution of categorical variables and are a way to visualize the information in frequency tables.
- For continuous numerical data, reporting the frequency of each unique entry is not an effective summary as many or most values are unique. Instead, a distribution function is required.
- The *cumulative distribution function (CDF)* is a function that reports the proportion of data below a value a for all values of a : $F(a) = \Pr(x \leq a)$.
- The proportion of observations between any two values a and b can be computed from the CDF as $F(b) - F(a)$.
- A *histogram* divides data into non-overlapping bins of the same size and plots the counts of number of values that fall in that interval.

Code



```
# load the dataset
library(dslabs)
data(heights)

# make a table of category proportions
prop.table(table(heights$sex))
```

Creating other graphs from the video

You will learn to create barplots and histograms in a later section of this course.

[Learn About Verified Certificates](#)

© All Rights Reserved

