



[Course](#) > [Section...](#) > [3.1 Su...](#) > [Group ...](#)

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](#)

Group By Group By

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

group then
summarize



RAFAEL IRIZARRY: A very common operation and data exploration

is to first split data into groups and then compute summaries for each group.

For example, we may want to compute the average and standard deviation for men

and women heights

Video



[Download video file](#)

Transcripts

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

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

Textbook link

This video corresponds to the [textbook section on grouping then summarizing](#).

Key points

- The `group_by()` function from **dplyr** converts a data frame to a grouped data frame, creating groups using one or more variables.
- `summarize()` and some other **dplyr** functions will behave differently on grouped data frames.
- Using `summarize()` on a grouped data frame computes the summary statistics for each of the separate groups.

Code

```
# libraries and data
library(tidyverse)
library(dslabs)
data(heights)
data(murders)

# compute separate average and standard deviation for male/female heights
heights %>%
  group_by(sex) %>%
  summarize(average = mean(height), standard_deviation = sd(height))

# compute median murder rate in 4 regions of country
murders <- murders %>%
  mutate(murder_rate = total/population * 100000)
murders %>%
  group_by(region) %>%
  summarize(median_rate = median(murder_rate))
```

Learn About Verified Certificates

© All Rights Reserved

