Cordero_week4.2

Joaquin Cordero

2024-07-01

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
scores <- read.csv("scores.csv")

head(scores)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 6    10   265 Regular
```

```
glimpse(scores)
```

```
## Rows: 38
## Columns: 3
## $ Count   <int> 10, 10, 20, 10, 10, 10, 10, 30, 10, 10, 20, 10, 10, 10, 10, 20~
## $ Score   <int> 200, 205, 235, 240, 250, 265, 275, 285, 295, 300, 300, 305, 30~
## $ Section <chr> "Sports", "Sports", "Sports", "Sports", "Sports", "Regular", "~
```

```r
scores_arrange <- scores %>% arrange(Score)

scores_duplicated <- scores_arrange[duplicated(scores_arrange), ]
print(scores_duplicated)
```

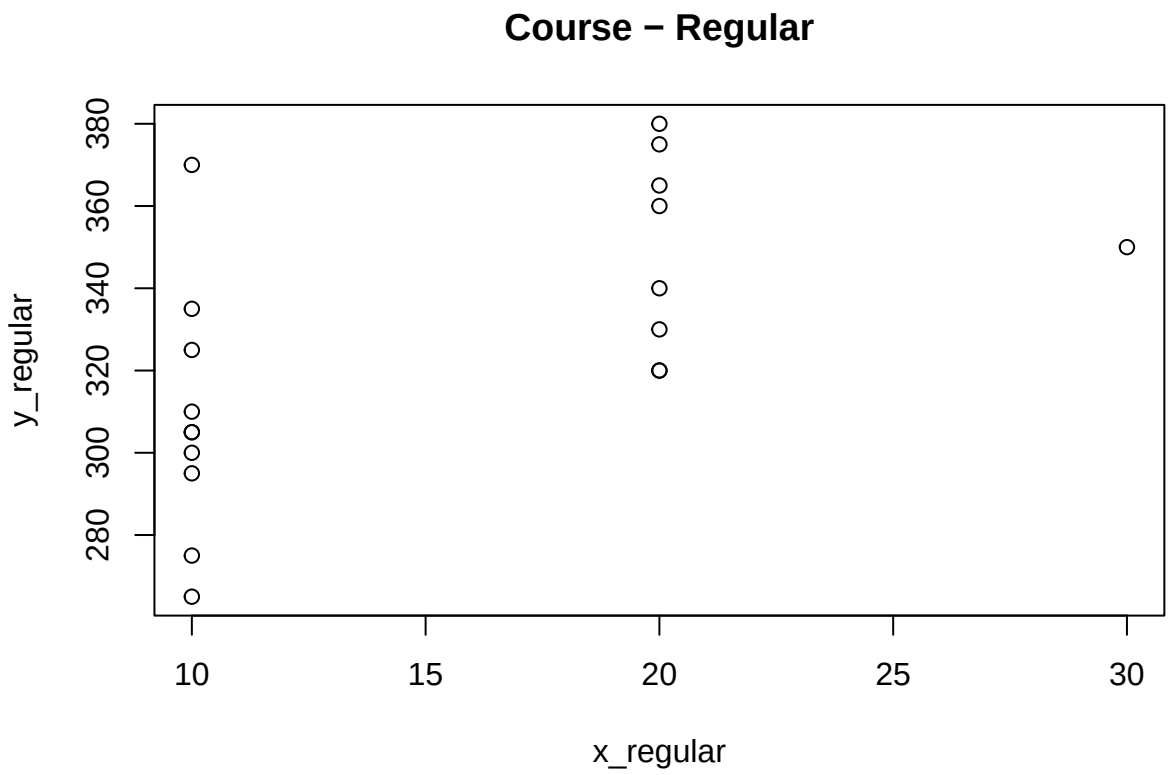
```
##    Count Score Section
## 14    10   305 Regular
## 19    20   320 Regular
```

```r
scores_unique <- scores_arrange[!duplicated(scores_arrange), ]
print(scores_unique)
```

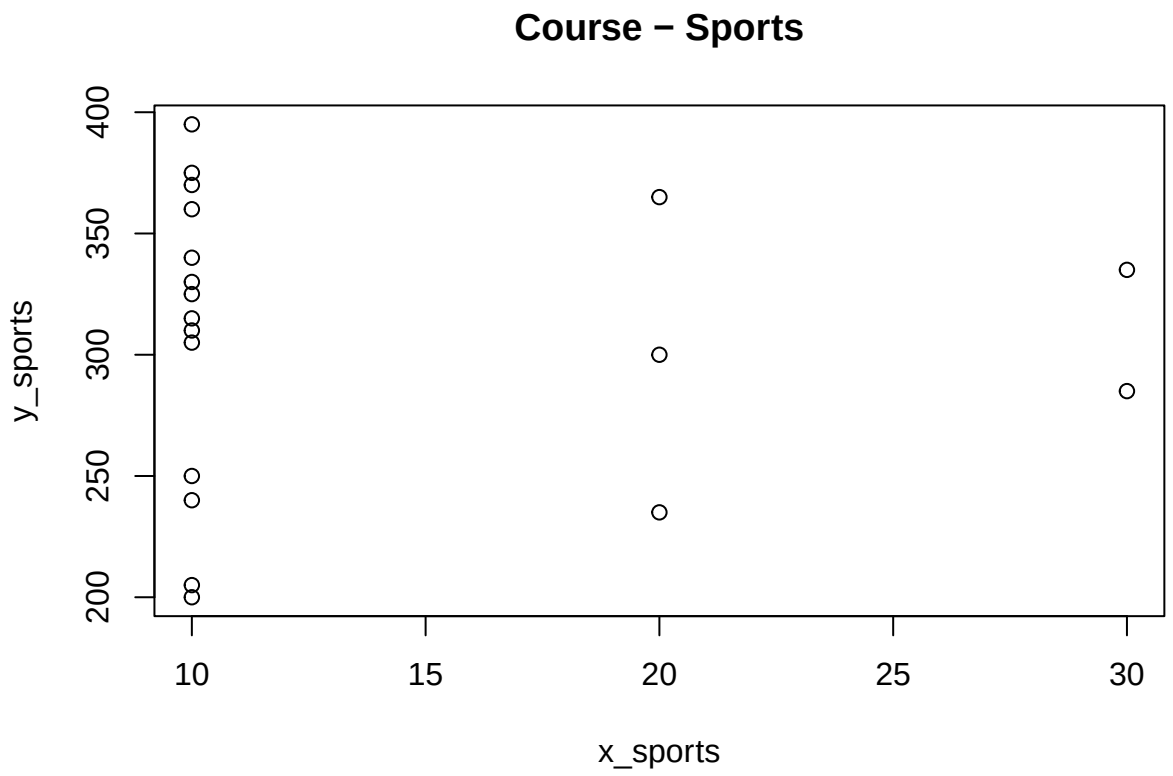
```
##    Count Score Section
## 1     10   200  Sports
## 2     10   205  Sports
## 3     20   235  Sports
## 4     10   240  Sports
## 5     10   250  Sports
## 6     10   265 Regular
## 7     10   275 Regular
## 8     30   285  Sports
## 9     10   295 Regular
## 10    10   300 Regular
## 11    20   300  Sports
## 12    10   305  Sports
## 13    10   305 Regular
## 15    10   310 Regular
## 16    10   310  Sports
## 17    10   315  Sports
## 18    20   320 Regular
## 20    10   325 Regular
## 21    10   325  Sports
## 22    20   330 Regular
## 23    10   330  Sports
## 24    30   335  Sports
## 25    10   335 Regular
## 26    20   340 Regular
## 27    10   340  Sports
## 28    30   350 Regular
## 29    20   360 Regular
## 30    10   360  Sports
## 31    20   365 Regular
## 32    20   365  Sports
## 33    10   370  Sports
## 34    10   370 Regular
## 35    20   375 Regular
## 36    10   375  Sports
## 37    20   380 Regular
## 38    10   395  Sports
```

```r
#1.What are the observational units in this study?
#-The observational units in this study are the two sections, the total grade
```

```r
#received, and the count of students that received that total grade.

#2.Identify the variables mentioned in the narrative paragraph and determine
#which are categorical and quantitative?
#-Categorical: the two sections (Sports or Regular)
#-Quantitative: Total Grade and course grade

#3.Create one variable to hold a subset of your data set that contains only the
#Regular Section and one variable for the Sports Section.

#4.
regular_section <- scores[scores$Section == "Regular", ]

x_regular <- regular_section$Count
y_regular <- regular_section$Score
plot(x_regular, y_regular, main = "Course - Regular")
```




```r
sports_section <- scores[scores$Section == "Sports", ]

x_sports <- sports_section$Count
y_sports <- sports_section$Score
plot(x_sports, y_sports, main = "Course - Sports")
```




```
#4a.Comparing and contrasting the point distributions between the two section,
#looking at both tendency and consistency: Can you say that one section tended
#to score more points than the other? Justify and explain your answer.
#-Looking at the point distributions between the two sections, it looks like the
#Regular course tended to score more points than Sports course. This is because
#more students in the 20 range per total grade earned scored higher than 320.

#4b.Did every student in one section score more points than every student in
#the other section? If not, explain what a statistical tendency
#means in this context.
#Yes, for the sports course students tend to score above 300. However, for the
#regular course students did not score more points than other students. In this
#context statistical tendency is the grade most students would receive.

#4c.What could be one additional variable that was not mentioned in the
#narrative that could be influencing the point distributions between the two
#sections?
#-A variable that could be influencing the point distributions between the two
#sections could be the prior knowledge student have. Some students would know
#more about sports than others and some would have no interest in sports at all.
```