Lab 4: Manipulate data with dplyr

Yasmine Pascal

02/11/2022

## Instructions

This document is the template you will use to type up your responses to the lab exercises that you can find on Canvas.

To complete the lab type your BRIEF answers and the R code in the spaces provided below, according to these guidelines:

* In general, if an exercise is numbered you will submit the R code that you used to answer the question in the gray area between the three ` marks.
* If you need to submit an answer in addition to the code type this between the line that starts #### and the beginning of the gray area for the code.
* Do not ever type the View() command in a lab report. This works only in RStudio on your computer, so don’t include that command in your lab report or you will not be able to compile the document, a process called *knitting*.
* To produce a document that you can submit on Canvas just click on the Knit button in the upper left quadrant of RStudio.
* We recommend pressing the Knit button after entering each answer: it’s very common to get errors, and they’re easier to fix if you can easily isolate the code that caused the error.
* The Knit button will create two files: one is a .Rmd file (R Markdown) and the other is an .docx file. You will need to submit both these documents for the lab assignment on Canvas.

## Preliminaries

#### Exercise 1

library(tidyverse)  
library(mosaic)  
library(ggformula)  
library(data1135)

## Preliminary data analysis

#### Exercise 2

This data was collected from January 1st 2016 to July 31st 2016.

summary(GoogleAnalyticsData)

## date channelGrouping deviceCategory sessions   
## Min. :2016-01-01 Length:5732 Length:5732 Min. : 1.0   
## 1st Qu.:2016-02-23 Class :character Class :character 1st Qu.: 41.0   
## Median :2016-04-16 Mode :character Mode :character Median : 153.0   
## Mean :2016-04-16 Mean : 335.1   
## 3rd Qu.:2016-06-08 3rd Qu.: 513.0   
## Max. :2016-07-31 Max. :11115.0   
## pageviews entrances bounces   
## Min. : 1.0 Min. : 1 Min. : 0.0   
## 1st Qu.: 87.0 1st Qu.: 41 1st Qu.: 24.0   
## Median : 339.0 Median : 152 Median : 74.0   
## Mean : 783.2 Mean : 333 Mean : 203.7   
## 3rd Qu.: 1064.0 3rd Qu.: 510 3rd Qu.: 233.0   
## Max. :18034.0 Max. :11063 Max. :7779.0

dim(GoogleAnalyticsData)

## [1] 5732 7

## Wringing insights from data

#### Exercise 3

GoogleAnalyticsData <- mutate(GoogleAnalyticsData, bounce\_rate = bounces / entrances)

#### Exercise 4

The proportion of rows in which the bounce rate is greater than 50% is .5898

prop( ~ (bounce\_rate > .5), data= GoogleAnalyticsData, success = TRUE)

## prop\_TRUE   
## 0.5898465

## Display Advertising

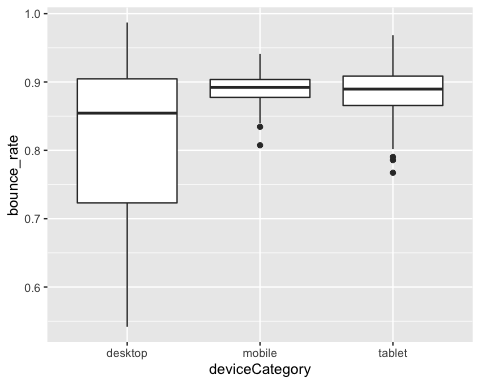
#### Exercise 5

displayAds <- filter(GoogleAnalyticsData, channelGrouping == "Display")

#### Exercise 6

Tablets have the greatest number of outliers, which means that they have the most observations which lie a considerable distance away from the data, which could infer that other factors are affecting tablet users. Desktop has the largest IQR which means that 50% of its data is spread out further compared to devices like mobile where 50% of its data is clustered tightly together.

gf\_boxplot(bounce\_rate ~ deviceCategory, data= displayAds)



## Display advertising vs email on mobile devices

#### Exercise 7

mobileDevs <- filter(GoogleAnalyticsData, deviceCategory == "mobile",channelGrouping == "Email" |channelGrouping == "Display")

#### Exercise 8

On average, display has a greater bounce rate than email, which means that users who receive an email are more likely to explore further than the first page on their website.

favstats( ~bounce\_rate | channelGrouping, data= mobileDevs)

## channelGrouping min Q1 median Q3 max mean  
## 1 Display 0.8074694 0.8774548 0.8920973 0.9036403 0.9410745 0.8899547  
## 2 Email 0.2000000 0.4883721 0.5641026 0.6250000 1.0000000 0.5615044  
## sd n missing  
## 1 0.02049832 213 0  
## 2 0.10515139 213 0