

Homework 02

DSO 545: Statistical Computing and Data Visualization

Fall 2018

Instructions

- Use R to answer all question. Write R code for each question.
- Submit your R file to blackboard with comments when needed
- USC won't tolerate any kind of cheating
- Good luck!

Case 01: Heavenly Chocolates

Heavenly Chocolates manufactures and sells quality chocolate products and its plant and retail store located in Saratoga Springs, New York. Two years ago, the company developed a website and began selling its products over the Internet. Website sales have exceeded the company's expectations, and management is now considering strategies to increase sales even further. To learn more about the website customers, a sample of 50 Heavenly transactions was selected from previous month's sales. Data showing the day of the week each transaction was made, the type of browser the customer used, the time spent on the website, the number of web pages viewed, and the amount spent by each of the 50 customers are contained in the files names `HeavenlyChocolates.csv`.

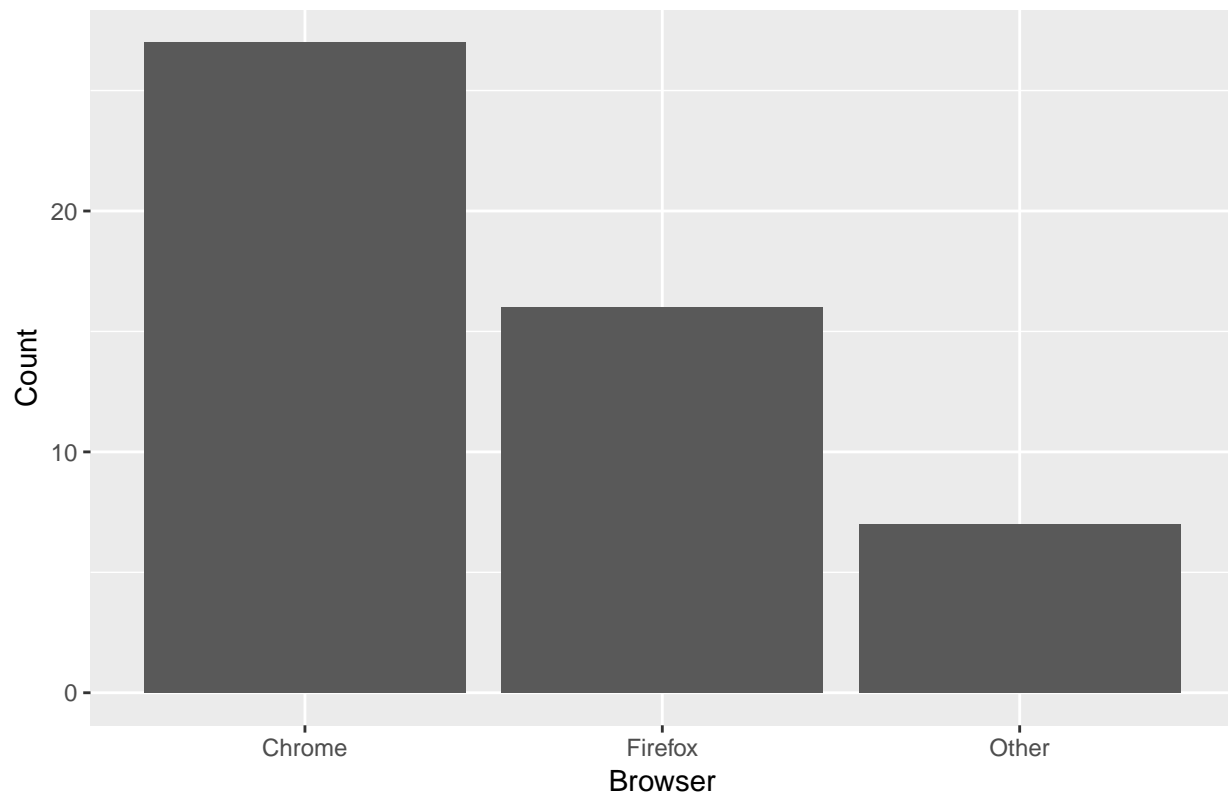
Heavenly Chocolates would like to use the sample data to determine whether online shoppers who spend more time and view more pages also spend more money during their visit to the website. The company would also like to investigate the effect that the day of the week and the type of browser have on sales.

Variables

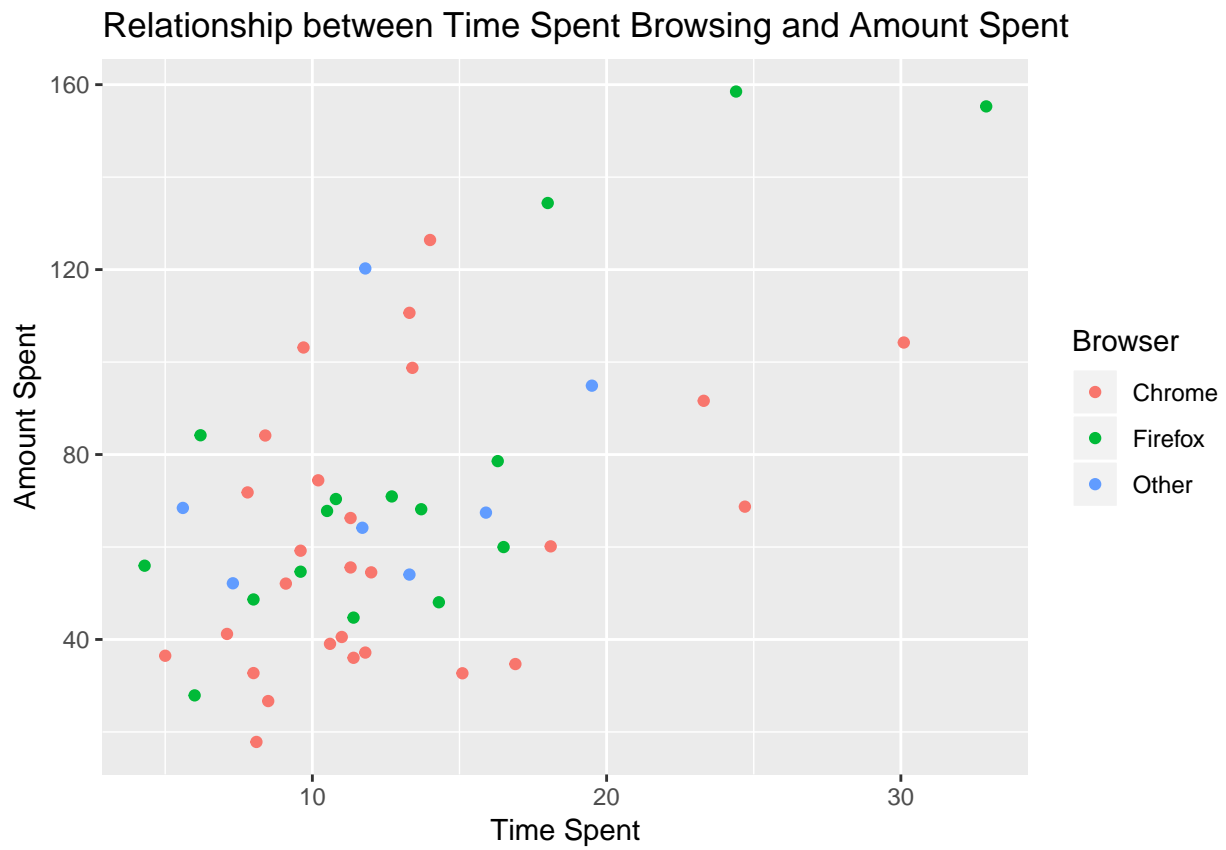
Variable	Explanation
Customer	customer's id
Day	the day of the week each transaction was made
Browser	the type of browser the customer used
Time	the time spent on the website (in minutes)
Pages_Viewed	the number of web pages viewed
Amount_Spent	the amount spent by each (in dollars)

1. (1 point) Create a duplicate of the following graph to which browser is used more by the Heavenly Chocolate website visitors?

Browsers Used to Visit Heavenly Chocolate Website

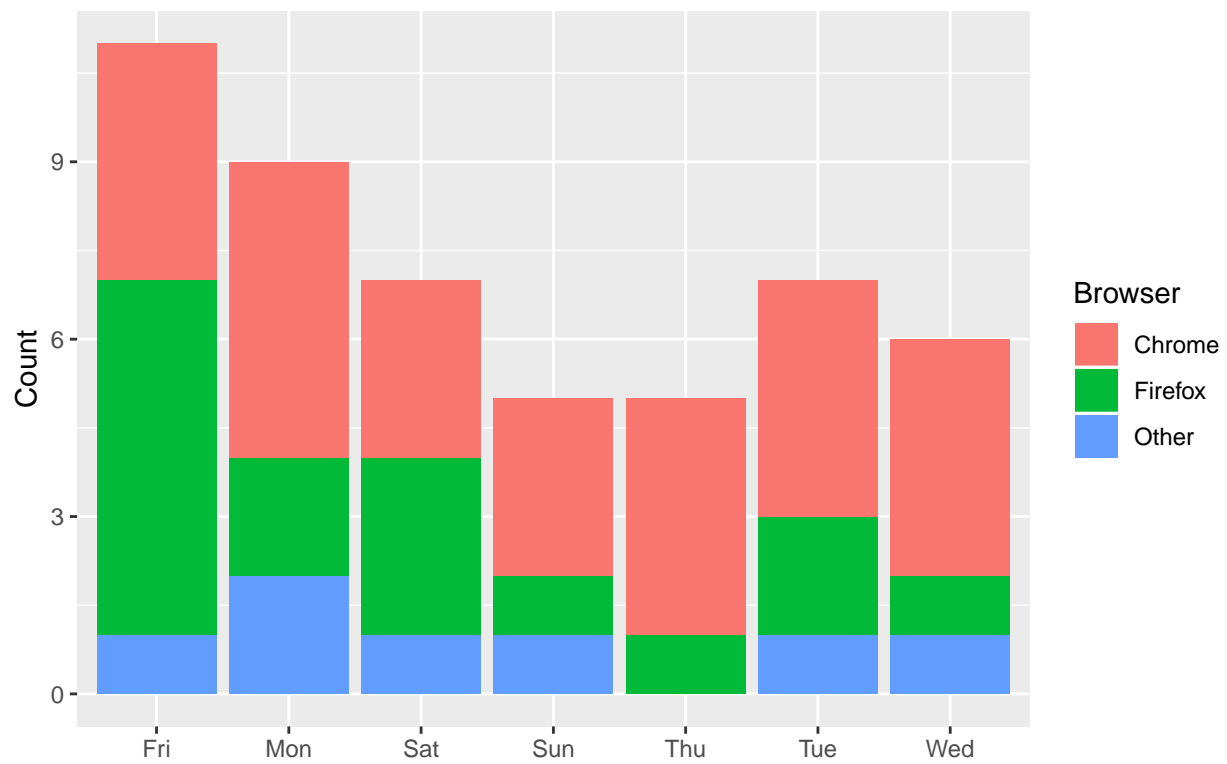


2. (1 point) Create an duplicate of the following scatterplot that shows the relationship between the time spent browsing, and the amount spent buying using different browsers.



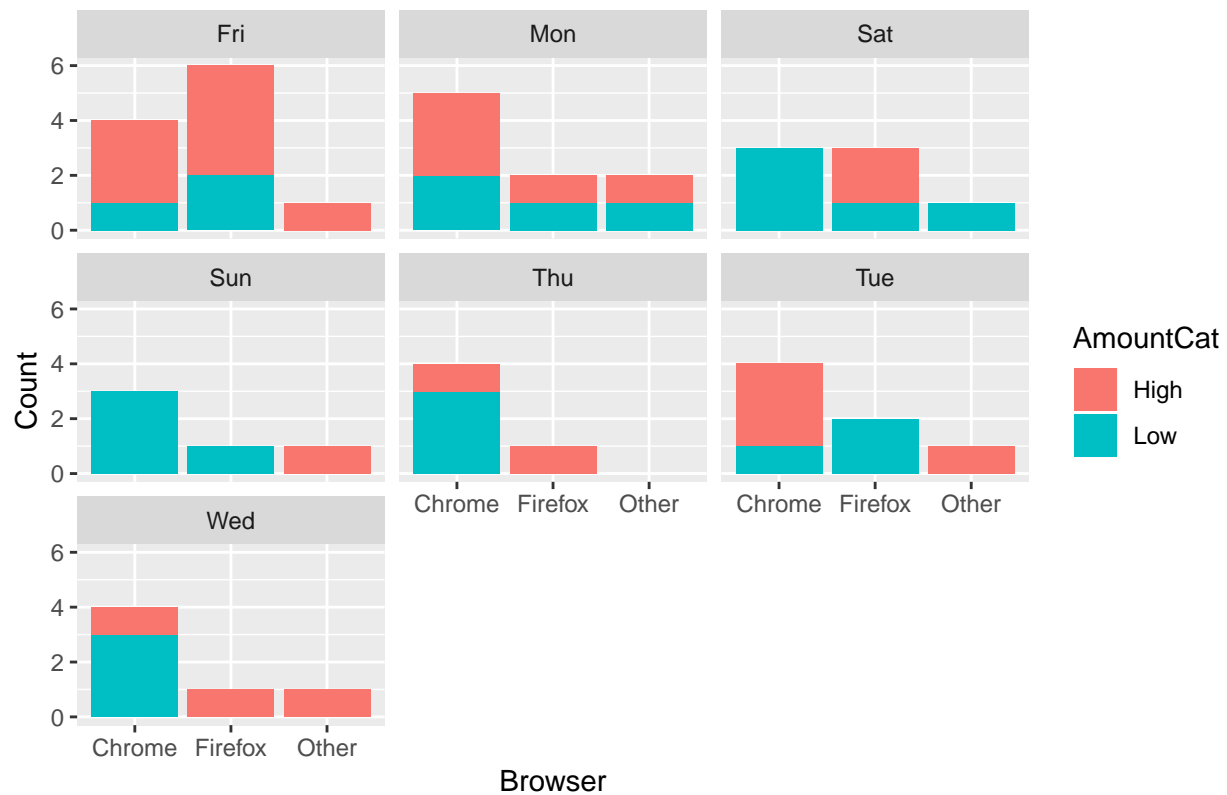
3. (1 point) Create a duplicate of the following graph to show the number of customers visiting the website on each day of the week, and what browser they are using.

Number of Website Visitors per Day per Browser

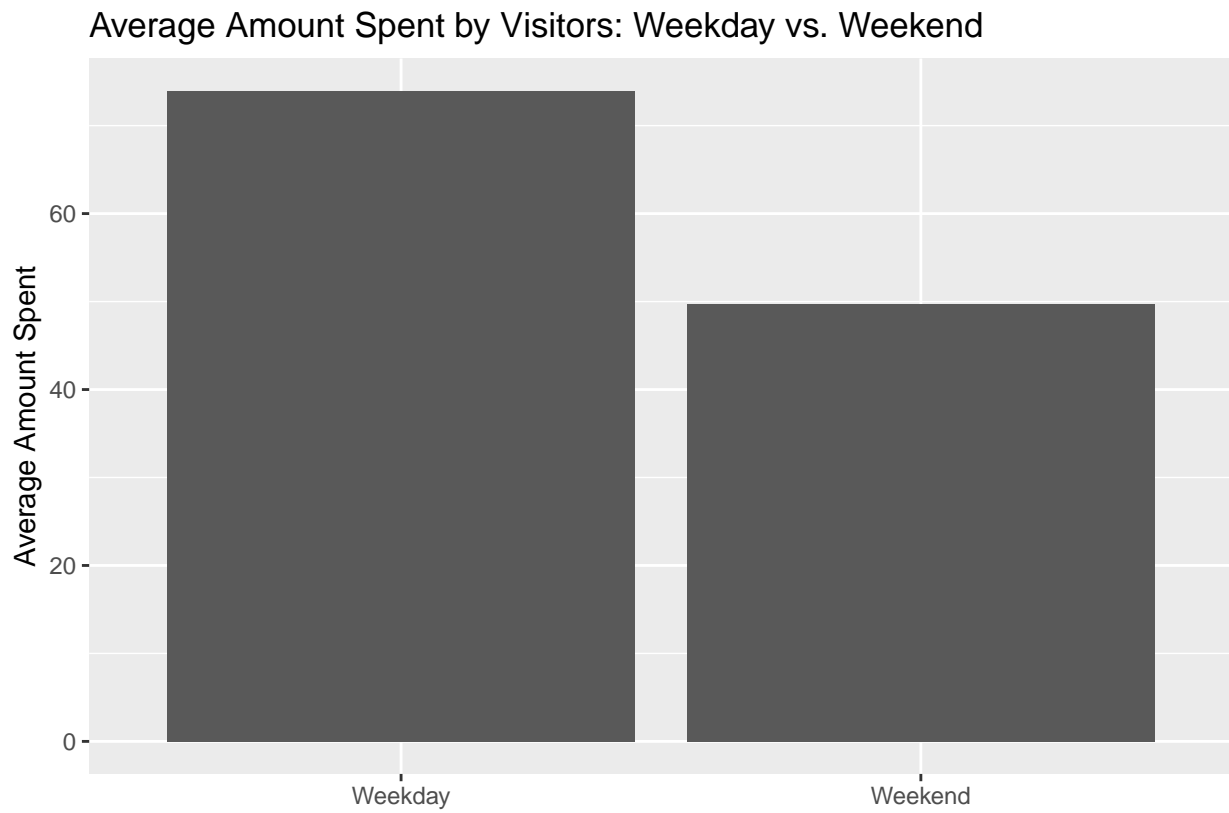


4. (1 point) Create a new variable called **AmountCat** as follows: If the amount spent is \$62 or below, then **AmountCat** is “Low”, otherwise, it is “High”.
5. (1 point) Create a duplicate of the following graph that shows the distribution of the **AmountCat** variable across different days of the week.

Distribution of ‘AmountCat’ across Different Days of the Week



6. (2.5 point) Create a duplicate of the following plot that shows you the average amount spent on a weekday (Mon, Tue, Wed, Thu, and Fri) vs. weekend (Sat and Sun).



Case 02: Which countries take the Syrian refugees?

According to CNN, the expanding Syrian refugee crisis highlights the differences among countries that welcome desperate migrants and those that don't.

Millions of Syrians are fleeing a homeland riven by more than four years of civil war. Some countries have taken in so many migrants it's caused a population spike, while others have done little or nothing at all. The dataset `syrian_refugees.csv` gives a country-by-country look at what is being done to address the worst refugee crisis ever.

Variables

Variable	Explanation
Country	Name of the hosting country
continent	Which continent is the country in?
refugees	Number of refugees in the country
Population	The population of the country

Questions

7. (2.5 point) Create a dot plot, showing the top 20 countries hosting the highest number of Syrian refugees.

