Project 1 – Chip's Chocolates

Chip Chesterfield started an online business 10 years ago selling homemade chocolates. Since then his business has become very successful and it has grown to the point where he needs a statistical consultant. You receive the following email from Chip:

Hello Mr./Ms. Consultant-

I'm not very good at math and I've never taken a statistics course and I need your help. We receive hundreds of orders per day and we're spending thousands of dollars on advertising so it is important for us to learn more about our customers. Attached to this email is a csv file with information from 554 randomly selected orders placed on our website last week. Here are some questions I hope you can help us with:

1
In general, how old are our customers?

2
In general, how do people hear about our website?

3
Is there is a relationship between the amount of money people spend and how they heard about our website? If so, what group(s) are spending more (or less) money?

4
I'd like to separate our customers into three groups (under 30 years old, greater than or equal to 30 and less than or equal to 45 years old, and over 45 years old). How many of the 554 customers belong to these three groups? Do these different age groups typically spend different amounts of money on their orders?

5
What percent of our customers are buying gifts for people? Are gift-buyers typically older or younger than people who aren't gift-buyers?

6
I heard a rumor that customers under the age of 26 who saw our advertisement on Facebook are spending less money per order than people who are not in this group. Is this true?

The data is in the file chip.csv

	A	B	C	D
1	money	age	hear	gift
2	38.70	30.0	friend	no
3	31.15	39.0	faceb	no
4	16.75	32.0	faceb	yes
5	11.44	22.0	faceb	no
6	14.78	50.0	google	no
7	46.77	29.0	faceb	no

Variables:

money (numeric)
The amount of **money** spent by the customer on the order.

age (numeric)
The **age** of the customer.

hear (categorical)
How did the customer hear about Chip's Chocolates?
faceb (learned about Chip's Chocolates from a facebook advertisement)
friend (learned about Chip's Chocolates from a friend)
google (discovered Chip's Chocolates during a google search)
magaz (learned about Chip's Chocolates from magazine advertisement)

gift (categorical)
Is the customer buying the chocolate as a gift for someone?
yes
no

For example, the first line of data in the csv file describes an order placed on the website by a customer who spent $38.70, is 30 years old, learned about Chip's Chocolates from a friend, and is not purchasing the chocolate as a gift.

Problem-by-problem Instructions:

1
Use **SAS**.
Describe the data set of customer ages. Include a histogram and report the mean, median, and standard deviation. Explain to Chip what the difference is between a mean and a median. Also compute a 95% confidence interval for the population mean age and explain to Chip how to interpret it.
(see SAS2 and SAS5 handout)

2
Use **R**.
Describe the categorial data set of the variable hear. Include a frequency table, pie chart, and bar chart.
(see R3 handout)

3
Use **R**.
Show Chip side-by-side boxplots comparing the typical amount of money spent by the four groups named faceb, friend, google, and magaz . If you believe there is a significant difference in spending behavior between any of the four groups then compare the values of their means and medians.
(see R3 handout)

4
Use **R**.
Create a function in R that identifies which of the three age groups each customer belongs to and use it to add another column to the data frame. Report the total number of customers in each group, then do a comparison similar to what you did for Question 3.
(see Lab2 and R3 handout)

5
Use **SAS**.
Provide side-by-side boxplots between the two groups (gift=yes and gift=no) for the variable age. Also compare the values of their sample sizes, means, and medians and provide comments.
(see SAS2 and SAS3 handouts)

6
Use **SAS**.
Provide Chip with side-by-side boxplots between the two groups (customers with an age less than 26 and who heard about Chip's Chocolates through facebook vs. customers who are not in this group). Also compare the values of their sample sizes, means, and medians and provide comments.
(see SAS2 and SAS3 handouts)

What should your project look like?

Part 1: Your report to Chip

Prepare a report for Chip that answers his six questions in order. Use Microsoft Word or a similar word processor with the individual graphs and plots copied/pasted in from SAS and R in the appropriate places. Do not leave large amounts of SAS and R output at the end of your answers to Chip's questions and repeatedly ask Chip to refer to it.

Part 2: Appendix

A1
Write a complete SAS program that reads in the chip.csv file and generates all the output you needed to complete Questions 1 and 5.

A2
Write a complete SAS program that reads in the chip.csv file and generates all the output you needed to complete Question 6.

A3
Write the R function you used to complete Question 4.

You may show me a preliminary version of your project if you would like some early feedback.