Customer Analytics Final Exam

General Instructions

This exam is made up of four questions. Whatever you do, **don't** answer all of them. Only answer the three you're most comfortable with. The three questions you answer will be weighted equally from a grading perspective.

You should be able to complete this exam in under three hours. However, you're welcome to spend as much time on it as you wish.

Refrain from communicating with anyone regarding this exam between now and Tuesday, April 24 at 3:15pm (the submission deadline). You're permitted to use your notes and browse the internet. However, you're not permitted to communicate with anyone (other than me and Google) about the exam. Doing so will constitute an honor violation.

Submission Guidelines

Once you've answered three questions to the best of your ability, please:

1. Save your answers in an R script with the following file-name structure "final-exam-computingid.R" (e.g., "final-exam-jpb2f.R").
2. Attach your submission to an email.
3. Make "2018 GCOM 7140 Final Exam Submission" the email subject.
4. Include an electronic signature and the following text in the body of the email, "*On my honor, I pledge that I have neither given nor received help on this examination. I worked alone on all aspects of it and the analysis I present is my own.*"
5. Send the email to me at `boichuk@virginia.edu`.

Late Policy

Early submissions are welcome. Late submissions will be penalized according to the following schedule:

- 0-2 hours late = -25%
- 2-24 hours late = -50%
- 24+ hours = no credit.

Getting Started

Start by creating an R script called "final-exam-computingid.R" (e.g., "final-exam-jpb2f.R"). At the beginning of this script, paste the following lines of code and run them to load the `completejourney`, `tidyverse`, and `gtrendsR` packages:

```
library(completejourney)
library(tidyverse)
library(gtrendsR)
```

In the remainder of your script, be sure to follow the Tidyverse Style Guide and to separate your answers with clear markers.

Follow-Up

I'll communicate your final exam grade to you no later than Friday, May 4. Your grades related to in-class participation, the homework assignments, and the group project will be made available to you on Blackboard by Friday, May 4 as well.

Thanks and Good Luck

Questions

Question 1

```
transactions %>%
  filter(between(quantity, 1, 20)) %>%
  mutate(
    retail_disc       = abs(retail_disc),
    coupon_disc       = abs(coupon_disc),
    coupon_match_disc = abs(coupon_match_disc),
    regular_price     = (sales_value + retail_disc + coupon_match_disc) /
                         quantity,
    loyalty_price     = (sales_value + coupon_match_disc) /
                         quantity,
    coupon_price      = (sales_value - coupon_disc) /
                         quantity,
    purchase_price    = case_when(coupon_disc > 0 ~ coupon_price,
                                  retail_disc > 0 ~ loyalty_price,
                                  TRUE            ~ regular_price),
    quantity_binned   = cut(quantity, breaks = seq(0, 20, 2))
  ) -> my_transaction_data
```

Using `my_transaction_data`, create two separate visualizations that depict the relationship between quantity purchased and purchase price with the functions `geom_boxplot()` and `geom_smooth()`, respectively.

a. Solely considering **the box plots you created in the first visualization**, craft one or two sentences to describe the relationship between quantity purchased and purchase price in the data.

b. Solely considering **the curve you created in the second visualization**, craft two or three sentences to describe the relationship between quantity purchased and purchase price in the data.

c. In four sentences or less, explain **why** the first and second visualizations are useful (or useless) depictions of the relationship between quantity purchased and purchase price in the data.

d. All things considered, craft two or three sentences to communicate how **you** would now describe the relationship between quantity purchased and purchase price in the data.

e. What marketing activities or principles of consumer psychology might help explain the relationship between quantity purchased and purchase price in the data? (Word limit: 75 words.)

Question 2

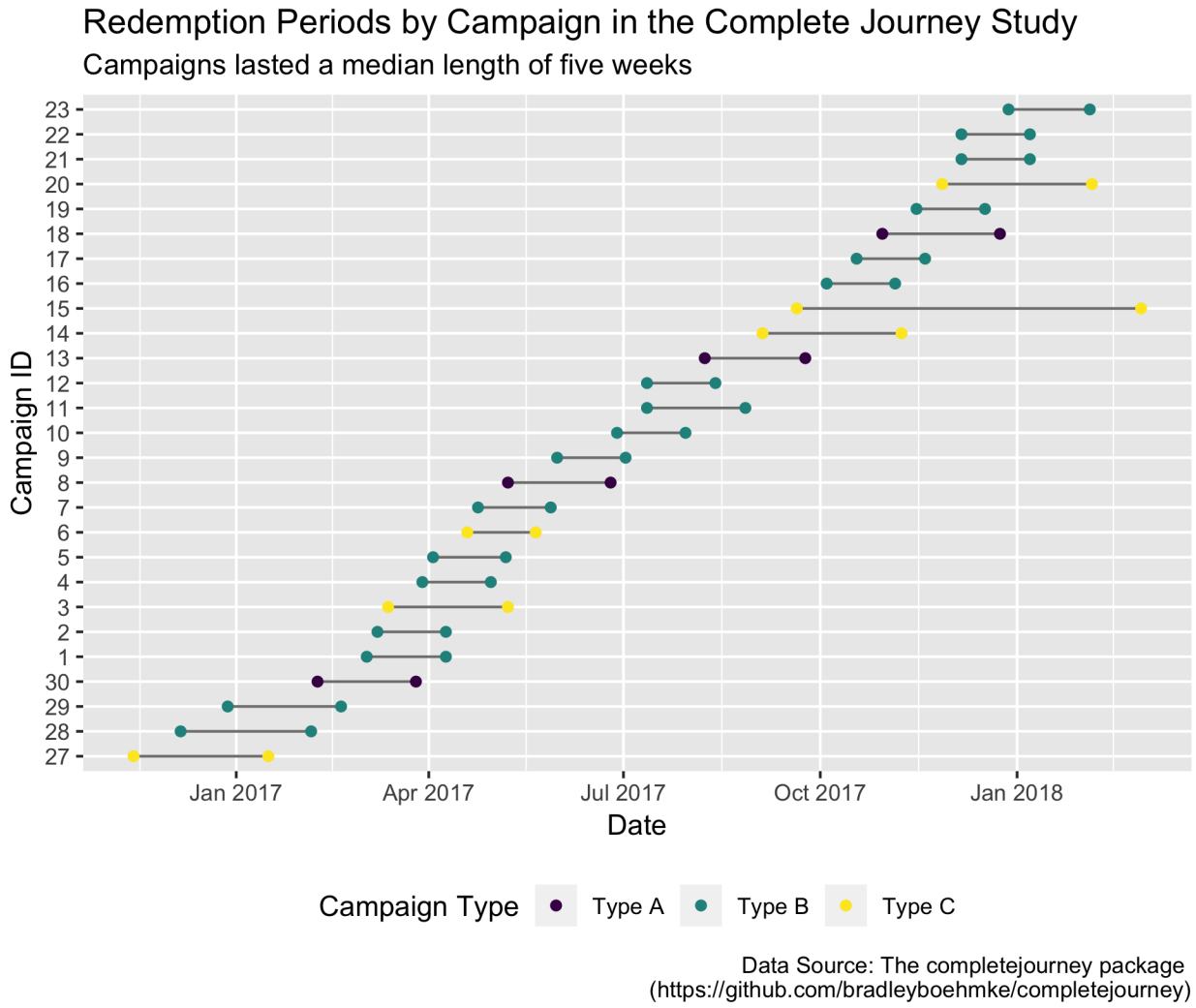
Using any of the datasets in the `completejourney` package, produce two exploratory graphics that force you to notice something revealing. Make your graphics with different geoms and different variables. For example, if you use `geom_histogram()` for your first graphic, use something other than `geom_histogram()` for your second. If you visualize `sales_value` from `transactions` in your first graphic, visualize something other than `sales_value` in your second.

Your graphics can either relate to one another or be unrelated. Either way, make sure you explain what your graphics display in 300 words or less. Your score on this question will depend on the degree to which your graphics produce accurate and interesting insights that you interpret correctly.

Question 3

Leveraging concepts from Chapter 12 of R for Data Science and Wickham (2014), explain why `campaign_descriptions` needs to be tidied. (Word limit: 75 words.)

Then tidy `campaign_descriptions` and recreate the following visualization:



Hint: This visualization shares several properties with Figure 12.2 in Tidy Data.

Adding the following layers to your visualization will:

1. specify the `limits` and `breaks` of the x axis,
2. move the legend to the bottom of the graph, and
3. make the visualization more self-explanatory.

```
scale_x_date(date_breaks = "3 months", date_labels = "%b %Y") +
theme(legend.position = "bottom") +
labs(
  title = "Redemption Periods by Campaign in the Complete Journey Study",
  subtitle = "Campaigns lasted a median length of five weeks",
  x = "Date",
  y = "Campaign ID",
  color = "Campaign Type",
  caption = "Data Source: The completejourney package
  (https://github.com/bradleyboehmke/completejourney)"
)
```

Question 4

Using the `gtrendsR` package, create an original, expository graphic that adheres to all of the techniques covered in Chapter 28 of R for Data Science. Full marks will be granted to graphics that:

- summarise their main findings with BLUF headings and/or subheadings,
- have meaningful labels and axis breaks,
- include captions that clearly and completely capture the data being analyzed,
- stem from scripts that use the grammar of graphics in a fluent manner, and
- force the reader to notice unexpected phenomena.