Final Exam Review

Coding and graphical questions

1. Describe a way or sketch out R code to find the mean of the cost vector below. Note that `mean(cost)` will give an error.

```
cost <- c("$1100", "$250.12", "$675")
```

2. Consider the following data frame:

```
msu.football <- data.frame(opponent = c("Washington State", "South Dakota State",
                                        "North Dakota", "Weber State", "Univ of Montana"),
                           points = c(0, 27, 49, 17, 134),
                           outcome = c("Loss", "Loss", "Win", "Loss", "Win"))
```

For each part below, explain what each line of code is doing (how each line of code helps produce the output). Then write the R output from the code below. Exactly one part will produce an error.

a.
```
for (i in 1:nrow(msu.football)){
  print(msu.football[i, 2])
}
```

b.
```
msu.football %>% filter(outcome == "Loss") %>%
  summarize(MaxPoints = max(points))
```

c.
```
msu.football %>% select(points) %>%
  group_by(outcome) %>%
  count()
```

d.
```
msu.football %>% ggplot(aes(x = points, y = outcome)) +
  geom_point() +
  labs(x = "Points Earned",
       y = "Outcome") +
  ggtitle("Points earned by outcome")
```

e.
```
msu.football %>% select(opponent) %>%
  mutate(CatGriz = case_when(
            opponent == "Univ of Montana" ~ "Yes",
            opponent != "Univ of Montana" ~ "No"
            )
          )
```

3. Describe a strategy to merge the two data frames defined below without losing any information (i.e., keep all rows from `df1` and all rows from `df1`), then write the output you'd expect to see.

```
df1 <- data.frame(school = c("MSU", "VT", "Mines", "Luther"),
                  state = c("MT", "VA", "CO", "IA"))
df2 <- data.frame(college = c("Mines", "MSU", "VT"),
                  enrollment = c(5794, 15688, 30598))
```

4. You would like to write a function that will take our original sample comprised of a single quantitative variable and create a bootstrap confidence interval for the true population mean, with a user-specified confidence level. Fill in the function below by adding R code or pseudocode wherever you see `<***>`.

```
bootstrap_means <- function(dat, conf.level = 0.95, reps = 1000) {
  # Function to generate a bootstrap distribution of means,
  # and calculate and report a confidence interval for the mean.
  # ARGS:
  #   dat = sample data (numerical vector)
  #   conf.level = confidence level as a decimal (defaults to 0.95)
  #   reps = number of bootstrap samples to generate (defaults to 1000)
  # RETURNS: a histogram plotting `reps` simulated means of size `sample_size`

  # Set up vector to store bootstrapped means

  <***>

  # Create bootstrap distribution of means
  for(i in 1:reps) {
    # Generate bootstrap sample

    <***>

    # Calculate and store mean of bootstrap sample

    <***>

  }

  # Use distribution of bootstrapped means to calculate confidence interval

  <***>

}
```

Free response questions

At least one of the following questions will be included on the in-class portion of the final exam. Answers will not be provided for this portion of the review worksheet.

1. Describe the server and UI components of code to create a shiny app.
2. Define data scraping (i.e., web scraping) and discuss why a basic understanding of HTML code is useful for data scraping.
3. Compare and contrast "supervised" and "unsupervised" statistical learning.
4. Compare and contrast "classification" and "clustering" algorithms. (Hint: Classification algorithms are an example of supervised learning; clustering algorithms are an example of unsupervised learning.)
5. Describe a situation where Git will automatically perform a merge when you pull from a repository.
6. Describe a situation where a merge conflict occurs when collaborating in GitHub.
7. Assume you write a function in R. What elements are necessary for documenting this function?
8. Describe at least two principles of good data visualization and include a sketch to demonstrate. (You will not be graded on your artistic ability, but keep it neat.)

Course evaluations

If you haven't yet, please take the time to fill out our course evaluation. We use course evaluations to reflect on what went well and what didn't go so well in a course, and to help us make changes that would improve the course in the future. We greatly value your feedback!