Assignment 0

Data 311: Machine Learning

Dr. Irene Vrbik

Objective

This assignment pertains to Lab 0, which serves as a refresher on R and an introduction to Quarto documents. The goal is to help students master the basics of creating dynamic and reproducible documents with Quarto, ensuring a solid foundation for the next assignment.

By the end of this assignment students should:

- Create a self-contained HTML document using Quarto, including updated metadata in the YAML header.
- Embed R code within the document and incorporate narrative text using R Markdown syntax.
- Complete basic R tasks to refresh fundamental skills.
- Embed images with cross-references and captions.
- Resize graphics produced by embedded R code.

Instructions

Complete the following tasks in a Quarto document (`assignment0_studentnum.qmd`). For this first assignment we will ask you to include both the Quarto file (`.qmd` source document) and the rendered output (HTML). For future assignments, only the HTML document will be required, with the source .qmd file requested only when issues arise that require troubleshooting or further clarification of your work.

Assignment Questions

Basic Quarto

1. (1 point) Create a Quarto document named `assignment0_xxx.qmd`, where `xxx` should be replaced with your student number.

2. (1 point) Update the YAML header of your Quarto document to include:

   - The assignment name as the main title,
   - The course name as a subtitle,
   - Your first and last name, along with your student number, as the author,
   - Enable the table of contents,
   - include the following option to ensure your images are embedded and viewable in your HTML document that will be uploaded to Canvas:

```yaml
format:
  html:
    embed-resources: true
```

3. (2 points) Create a level-one header titled **Introduction**. Below this header, write a brief introduction about yourself (including your name, year, and major). Use bold and italics formatting on some words in your paragraph.

4. (2 points) **:** Below your introduction, include an image that represents something about you (e.g., a photo of yourself, your pet, or a place you like to visit). Add a caption to the image. For example:

Figure 1: This picture was taken after a fun game of mixed doubles tennis with my husband in West Kelowna.

Data Wrangling

Unless otherwise specified be sure to display the R code along with the output in the rendered document (`echo: true` should be the default anyways).

5. (2 points) Download the data CSV file from https://irene.vrbik.ok.ubc.ca/data/sales_data.csv and save it in a subdirectory within your assignment folder (refer to Lab 0 for recommended folder setup). Load the dataset into R using a relative path and save it to an R object (naming of your choice).
6. (1 point) Display a preview of your data by using the `head()` and `str()` functions on your data frame. For (1 bonus point), modify your YAML to change how your data frame is printed by following the instructions here: https://quarto.org/docs/computations/r.html#data-frames (I use the `paged` option for lecture slides).
7. (1 point) Identify and remove any duplicate rows in the dataset using appropriate R functions.
8. (1 point) Check for missing values in the dataset.
9. (1 point) Filter the dataset to display only the sales made in New York.
10. (1 point) Identify the date on which the highest total sale was recorded in New York.
11. (1 point) Determine the most frequently used payment method by customers.

Data Visualization

12. (2 points) Create a histogram of Customer Age and label the x-axis as "Customer Age". Adjust the plot dimensions so that it is taller than it is wide.
13. (2 points) Create a scatter plot with `Quantity Sold` on the x-axis and `Price` on the y-axis. Ensure that all axis labels are clear and readable. Title the plot "Relationship between Quantity and Price" and label the chunk using the naming convention that facilitates cross-referencing.
14. (2 points) Based on the scatter plot from the previous question, describe any apparent relationship between `Price` and `Quantity Sold`. Include a cross-reference to the figure number in your response.
15. (1 point) Submit both your rendered HTML and the `.qmd` source file to Canvas for grading. Verify that your HTML document displays as expected before submission.

Total points: 21 + 1 bonus points