DATA423-20S1 - Data Science in Industry

Assignment 1

In brief:

Part 1:

Create a Shiny app using **RStudio**. Using the supplied comma separated variables (csv) file load the data and use Shiny to s
and visualise the information within. There is no need to perform any modelling.

The deliverable should be a pair of files: ui.R & server.R that I should be able to run and grade.

Part2:

Using the supplied data write a short document. In it explain what you saw in the plots that was special / unusual / curious ab
data set. Try to say something about each plot.

This deliverable should be a pdf document.

This assignment will demonstrate

a. your grasp of Shiny (I am not looking for advanced Shiny skills here)
b. your use of reactive inputs
c. your ability to self-discover aspects of R esp. how to summarise / visualise data sets
d. your ability to interpret the information
e. your ability to communicate the relevant information clearly

In Full:

Steps:

1. Create a shell of Shiny app. Plan whether you want a *sidepanel/main* layout or a *fluidpage* layout or something more ambiti
   Design how the user would progressively get more information by interacting with the page.

2. Add your name to the title part of the ui so it is clear to see whose app is running. Something like:

```
...
title = "Assignment 1 - John Smith",
...
```

3. Place the CSV file in the same location as the ui.R, server.R files. Load the csv file using something like:

```
dat <- read.csv("data.csv", header = TRUE)
```

4. Provide several visualisations about the data set. As a minimum provide the following a) a **vcd::mosaic** plot of discrete var
   b) a **GGally::ggpairs** plot, c) **corrgram::corrgram** correlation chart, d) **visdat::vis_miss** showing missing data, e) a
   DT::datatable a listing of the data, f) a **graphics::boxplot** that responds to a *slider* that controls the threshold for markin
   as outliers and checkboxes that cause the data to be centered and scaled, g) a **rising-value chart** would help to detect ga
   continuous data.

5. Test your shiny app to ensure it will run on another machine, (use *Session / Restart R* and *Session / Clear Workspace* in the menu). Every year students supply assignments that will not run without the marker making changes - typically they forget declare packages (i.e. library(tabplot)) or use variables that are set up outside of the supplied code. This will lose you poin

6. Write up what you have learned about features of the data set. Convert this document to a pdf.

7. Use the course LEARN website to upload the **pdf**, **ui.R** & **server.R** files.

Here are some things to consider:

Do the data types of any variables need to be programmatically altered? Are some strings actually dates? Are factors really order factors? Are there any cyclic variables?

Provide some summary statistics about the data set. **summary()** or **tibble::glimpse()** can achieve this. In particular make sure t number of variables and the number of observations are presented in your app.

What column identifies observations? Use this column appropriately in the visualisations.

When doing things like scaling: **data2 <- scale(data, center = input$center, scale = input$scale)** This will not work on non- data. You will need to only scale the subset of columns of the data set that are numeric.

Marking:

This assignment is worth 20% of your final grade

I am looking for charts that reveal features of this data. It has been synthetically constructed to have interesting features - can yo them all?

I am looking for charts that operate selectively on only those column types for which it is appropriate. For example *box plots* that plot the factor columns and *scatter plots* that do not plot the factor columns. Mosaic plots that only use discrete columns.

I am looking for charts that change as the user interacts with checkboxes and sliders.

I will reserve extra marks for something impressive in your shiny app although I will not know what that is until I see it. Go on, imp

The text document should show consistent and correct thinking about outliers, missing data, and homogeneity.

Ass1Data.csv 24 February 2020, 11:13 AM

Submission status

Attempt number	This is attempt 1.
Submission status	No attempt
Grading status	Not graded
Due date	Friday, 6 March 2020, 6:00 PM
Time remaining	10 days 1 hour
Last modified	-
Submission comments	

▸ Comments (0)

Add submission

You have not made a submission yet.

◀ Course outline

Jump to...

Introd